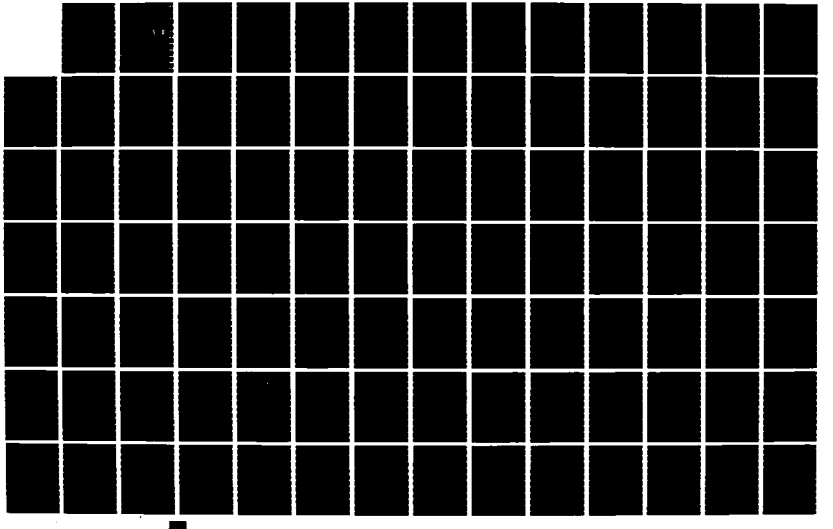
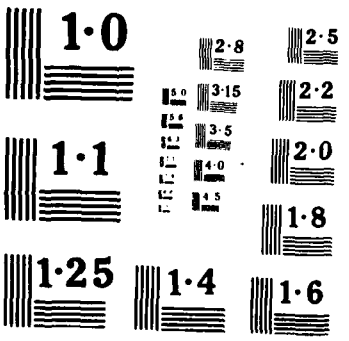
AD-A170 125 DETECTION ESTIMATION AND MULTIDIMENSIONAL PROCESSING OF 1/2
SINGLE EVOKED POT..(U) PURDUE UNIV LAFAYETTE IN EEG
SIGNAL PROCESSING LAB J I AUNON ET AL. SEP 85
UNCLASSIFIED AFOSR-TR-86-0367 F49620-83-K-0031 F/G 6/16 NL




1·0
1·1
1·25
1·4
1·6
1·8
2·0
2·2
2·5
2·8
3·15
3·5
4·0
4·5

Detection, Estimation, and Multidimensional Processing of Single Evoked Potentials

C. D. McGillem
J. I. Aunon

EEG Signal Processing Laboratory
School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907

AFOSR-F49620-83-K-0031
August 1985

Final Report for Period 1 April 1983 to 31 March 1985

Prepared for
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Bolling Air Force Base, DC 20332

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1 REPORT NUMBER	2 GOVT ACCESSION NO. ADA170 125	3 RECIPIENT'S CATALOG NUMBER
4 TITLE (and Subtitle) Detection, Estimation and Multidimensional Processing of Single Evoked Potentials		5 TYPE OF REPORT & PERIOD COVERED Final: 1 May 1983 30 April 1986
		6 PERFORMING ORG. REPORT NUMBER AFOSR-TR. 86-0367
7 AUTHOR(s) J. I. Aunon C. D. McGillem		8 CONTRACT OR GRANT NUMBER(s) F49620-83-K-0031
9 PERFORMING ORGANIZATION NAME AND ADDRESS School of Electrical Engineering Purdue University West Lafayette, IN 47907		10 PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2313/A4
11 CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NL Bolling AFB, DC 20332-6448		12 REPORT DATE September 1985
		13 NUMBER OF PAGES 144
14 MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15 SECURITY CLASS. (of this report) Unclassified
		15a DECLASSIFICATION DOWNGRADING SCHEDULE

16 DISTRIBUTION STATEMENT (of this Report)
Approved for public release. Distribution unlimited.

17 DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18 SUPPLEMENTARY NOTES

19 KEY WORDS (Continue on reverse side if necessary and identify by block number)
Evoked Potential, event related potential, pattern recognition, time-varying filter, feature selection, classification, artifact

20 ABSTRACT (Continue on reverse side if necessary and identify by block number)
A new method of generating and selecting features for computer classification of ERP waveforms is described. The technique employs features that are time samples of measured ERP waveforms that have been bandpass filtered. An improvement of more conventional techniques is shown with both simulated and measured data. A comparison of performance of a quadratic classifier with optimum feature selection is made with one using (suboptimal) forward sequential feature selection. Details of the design and performance
(continued on back)

DD FORM 1 JAN 73 1473

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

are presented for an optimum time-varying filter for estimating ERP waveforms. The filter accepts inputs from multiple electrodes and generates an estimate for the waveform at a single electrode. Results for both simulated and measured data are presented. Some preliminary studies of artifact generation in ERP waveforms are described.

DETECTION, ESTIMATION, AND MULTIDIMENSIONAL PROCESSING OF SINGLE EVOKED POTENTIALS

C.D. McGillem
J.I. Aunon

EEG Signal Processing Laboratory
School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907

August 1985

Final Report for Period 1 April 1983 to 31 March 1985

Prepared for
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Bolling Air Force Base, DC 20332

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

This report describes work carried out under AFSOR Contract No. F83K0031 entitled "DETECTION, ESTIMATION, AND MULTIDIMENSIONAL PROCESSING OF SINGLE EVOKED POTENTIALS." The overall objective of this research has been to develop new methods of processing single event related potentials (ERPs) to allow more effective information extraction to be carried out. Considerable progress has been made in developing methods for detecting the occurrence of ERPs, for distinguishing between ERPs produced by different stimuli, for separating the ERP waveforms from the ongoing electroencephalogram (EEG), and in obtaining a better understanding of the complexities of the ERP waveforms themselves. In Section 2 of this report, a new type of classifier is described that gives very good performance in distinguishing between ERPs generated by different stimuli. The classifier makes use of time samples of bandpass filtered ERP waveforms. Very high classification accuracies were obtained with this technique.

Section 3 presents the results of a preliminary study to compare the accuracy of classification using a quadratic discriminant function when optimally selected features are used as compared with performance when suboptimal features are selected by the forward sequential feature selection procedure. Among the conclusions of this study was the fact that although performance with the optimal features was somewhat better, it was not enough to warrant the extra computation required for their determination.

Section 4 describes a very effective filter for separating the ERP from the ongoing EEG. The filter is a multidimensional, time-varying linear operator that makes use of measured or estimated statistical properties of the ongoing EEG or ERP waveforms. Tests on both simulated and measured data show very high rejection of the ongoing EEG and retention of the ERP when data from two channels (i.e., two electrode sites) are used. Once the filter is designed, it is very easy to use and could be used on-line if desired.

Section 5 describes some preliminary studies on modeling the ERP waveform. It is shown in these studies that visual inspection of an averaged ERP waveform cannot be relied upon to differentiate between true components

of the waveform and artifacts corresponding to valleys or ridges produced by the interaction of adjacent components. Analysis of the measured latency variations of peaks in a given neighborhood appears to provide a method of distinguishing between true components and artifacts.

Section 6 provides a description of a new data acquisition and processing system currently in use in the EEG Signal Processing Laboratory. An IBM Personal Computer now runs experiments, collects data, and performs signal processing tasks. The system is both powerful and flexible.

The remaining sections list the professional personnel and summarize the awarded degrees and publications which have resulted from this contract.

2. IMPROVED CLASSIFICATION TECHNIQUES FOR ERP WAVEFORMS

2.1. Introduction

The focus of this research was to obtain a better method of feature transformation. A transformation procedure producing higher classification accuracies has the potential of bettering our understanding of ways to extract basic information contained in ERP data because the transformed features may be more representative of the information bearing components within the data. This introduction serves as a review of the statistical pattern recognition problem.

Statistical pattern recognition has been applied to distinguish features in measured waveforms that may be used to reliably detect or classify an event related potential (ERP) generated by the brain in response to a different stimuli (Donchin 1975). The results obtained by using these statistical tools can increase understanding of the EEG components, event related potentials, and ultimately brain function.

Pattern recognition may be thought of as a system involving: data measurement and recording; data transformation; feature selection; and classification. A feature is some measured value from the data or a value that is derived from measured values of the data. It is considered a random variable and is used as an input to the statistical classifier. In EEG signal processing, the data that are measured are the amplitudes of the voltages measured between pairs of electrodes attached to the scalp. These voltages are amplified, filtered, sampled, quantized, and recorded by a digital measurement system.

The most straightforward procedure for applying statistical pattern recognition techniques to ERP waveforms is to use the amplitudes of the signal at the sampling instants as the features (McGillem 1981, Donchin 1975, Aunon 1982a, Aunon 1982b, Sencaj 1979, Vidal 1977, Childers 1966, and Moser 1982). The signal is usually lowpass filtered and then resampled at a lower rate to reduce the number of possible features. Other approaches involve feature transformations such as frequency domain analysis (Moser 1982), principal component analysis (Van Hoek 1974), factor analysis (John 1973), or the

Karhunen-Loeve transform (Fukunaga 1970) to map the data into orthogonal components.

Features to be used by the classifier can be based on a variety of criteria (Mucciardi 1971 and Kanal 1974). The stepwise linear discriminant analysis (SLDA) program, based on a stepwise feature selection method (Kanal 1974) available in the BMD-07 (Dixon 1975) computer program package, is a suboptimal feature selection and classifier program that has been applied to EP recognition (McGillem 1981, Donchin 1975, Aunon 1982b, Vidal 1977, and Donchin 1966). Suboptimal feature selection techniques will not in general produce the subset of features that gives the best discrimination between the classes, but they greatly reduce computational requirements over optimal feature selection techniques. In this study forward sequential feature selection (Mucciardi 1971) was the method employed.

The EEG has been modeled as a nonstationary random process. Researchers have segmented the EEG into short time sections and carried out frequency analyses (Jansen 1981a, Jansen 1981b, and Sanderson 1980). The EEG is considered to have a time-varying spectrum and this type of analysis attempts to measure the underlying spectrum for particular portions of the data. Frequency spectra computed for different data segments showed significantly different underlying spectra.

Other work has focused on the computation of the time-varying spectra of the EEG recorded from subjects as light-flash stimuli were applied (Aunon 1977). In this study the data were windowed with a short rectangular window and the spectrum was computed. The window was then moved in time and a new spectrum was computed. Plotting and observing these spectra in a 3-dimensional form graphically illustrated a time-varying magnitude of certain frequencies. Figure 2-1 portrays an example of this type of plot, in which the magnitudes are plotted for points on the time-frequency plane. This shows the time-varying property of the frequency components. Other work has focused on the computation of the time-varying spectrum by the maximum entropy method (Pomalaza 1979).

The measured data d(t) of the ERP is considered to be composed of the evoked signal which is composed of deterministic components with certain randomly varying parameters, the ongoing EEG which is an additive random process consisting of the superposition of components similar to those of the signal, plus an independent additive noise:

$$d(t) = s(t) + e(t) + n(t) \qquad (2\text{-}1)$$

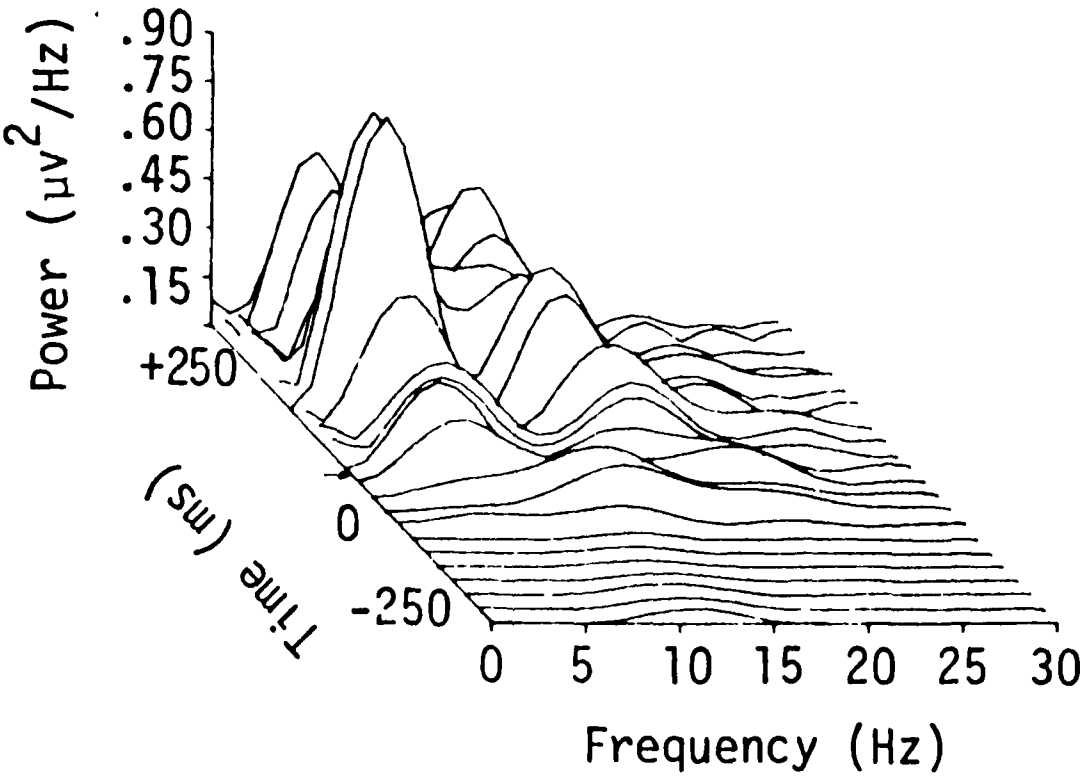


Figure 2-1 An Example Plot of a Time-Varying Spectrum for an Averaged Evoked Potential. Taken from Aunon (1977).

where s(t) = EP signal,
e(t) = ongoing EEG,
and n(t) = noise.

The components of the EP signal are considered to be generated in a similar manner for repeated stimulations.

The noise term n(t) accounts for the lumped instrumentation and measurement noise which should be independent of s(t) or e(t). Included in this term are muscle artifactual noise which is a time-varying process. Attempts are made to eliminate records contaminated with artifactual noise so that the noise term may be considered to contain only noise generated by an independent stationary random process.

In the feature transformation method considered here the time-varying spectrum underlying the signal is exploited to improve classification and detection accuracy as measured by classification error bounds and to identify the underlying components and their parameters. This is similar to methods proposed for use in speech processing (Tanaka 1979) and in noise pollution source recognition (Moukas 1982). Estimation of the frequency components over short time intervals will help identify the time-varying components at the various times. If a particular frequency or band of frequencies is found to be prevalent or reliably detected in the same small time segments of signals from an ensemble of data, that portion may be considered to be mapped into the signal space corresponding to a short duration sinusoidal component.

The present study tested the efficacy of using selected amplitudes from the time-frequency plane or, equivalently, portions of the real part of the time-varying spectrum, as features in classifiers to distinguish brain evoked potentials. In an early study features were selected from both the time and frequency dimensions, but the geometrical relationship between these dimensions was not considered to form the two-dimensional space (Moser 1982). The amplitudes of the time and frequency components were not taken in pairs but singly, hence the time varying nature of the spectrum was not taken into account.

Amplitudes at particular times, or at particular latencies of the evoked potential may contain energy from a wide range of frequencies. In classification studies the ERP is usually low-pass filtered before sampling (Aunon 1982b). The pass band might typically be 0.1 Hz to 25 Hz. If the amplitude of a signal at a particular time was measured after the signal had been filtered to allow only a narrow band of frequencies to pass, this would approximate the amplitude of a region in the time-frequency plane. Such amplitudes can be

used as features for classification and detection of single evoked potentials. A two-step feature selection/classification procedure as shown in Figure 2-2 was developed and tested to investigate the effectiveness of features taken from the time-frequency plane. The method shown in Figure 2-2 greatly reduces the computational requirements over selecting from all possible features of the time-frequency plane in one step. The two-step classification process first transforms the windowed signals into frequency components, then it selects the frequency features giving the best classification performance using the training set signals. These selected frequencies are then used to set the center frequencies of a set of band-pass filters which process the raw data. The amplitudes at the outputs of these filters at various times are then used as a set of features from which the final feature selection routine chooses the features that will be used by the classifier. This process is described in detail in Section 2.3.

Artificially generated EEG data and actual EEG data were used to test the procedure as described in Section 2.4. Data from four subjects participating in visual stimulation evoked potential experiments were tested and the results analyzed. Several sets of artificial data were generated and tested. They were composed of various types of signals added to either real EEG data or computer generated random noise data with various signal to noise ratios. The classification and detection accuracies are compared with those attained by more conventional methods and show a significant improvement in most cases.

2.2. Classifier and Feature Selection

The selection of the subset of features to be used from the complete feature set and the design of the classifier which uses the selected features is of central importance for obtaining high classification accuracy. The design of the classifier and feature selection process are discussed together in this section because the method chosen for feature selection is connected with the design of the classifier and the results of classifying the data records of the training set. The design of the classifier is discussed followed by a description of the procedure for estimation of the necessary statistics from the data. Next the error bound computation is described, followed by the feature selection process. The criteria for feature selection are based on the error bounds computed by the classifier from the training data set. Lastly, the classifier and feature selection algorithm applied to the detection problem is discussed.

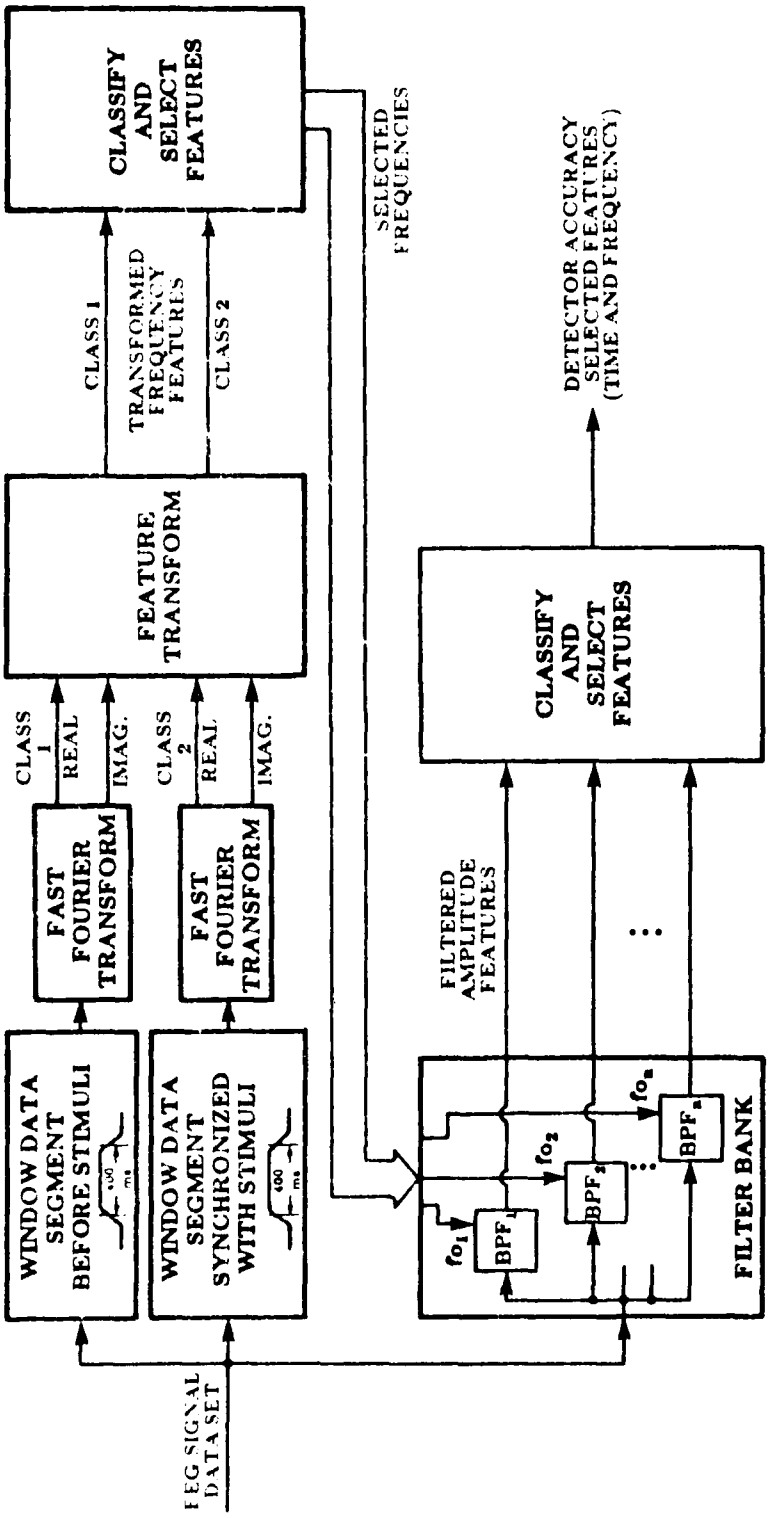


Figure 2-2 Block Diagram of 2-Step Classification/Feature Selection System, Setup for Detection

Under the assumption that the features are jointly Gaussian random variables the log-likelihood ratio classifier for the 2 class case is:

$$(\mathbf{x} - \mathbf{m}_1)^T \sum_1^{-1} (\mathbf{x} - \mathbf{m}_1) - (\mathbf{x} - \mathbf{m}_2)^T \sum_2^{-1} (\mathbf{x} - \mathbf{m}_2) \qquad (2\text{-}2)$$

$$+ \ln \frac{|\sum_1|}{|\sum_2|} \gtrless T,$$

where $\mathbf{x}$ is the sample feature vector to be classified,

$\mathbf{m}_i$ is the mean feature vector of class i,

$\sum_i$ is the covariance matrix of class i,

$\sum_i^{-1}$ is the inverse covariance matrix of class i,

and T is the decision threshold to which the function is compared.

If the function is less than T, the sample is classified as class 1; if it is larger than T, then it is classified as class 2. This classifier minimizes the total probability of error if it is used as a Bayes classifier where $T = 2 \ln \frac{P_2}{P_1}$, where P_i is the probability of occurrence of class *i*. Various other probability of error criteria could be used which would change the decision threshold value of the classifier (Fukunaga 1972). An advantage of this classifier is that it works well even if the true probability density is not exactly Gaussian, but has a Gaussian like shape (Kazakos 1982a and 1982b). In addition, it is a relatively easy function to compute and conducive to a computerized on-line real-time detector once the features have been selected.

In the detection problem, one class is the EEG with no signal or ERP and the other is the EEG with a signal. The threshold is then changed to adjust the ratio of the number of false detections to the number of correct detections. This is a convenient parameter for changing the detector function because it may not be known *a priori* what the probability of occurrence of the signal is, and the threshold value can easily be altered to adjust detection performance for a particular data set.

This quadratic classifier becomes a linear classifier if it can be assumed that both classes have the same covariance matrix. The quadratic classifier is used here so that assumption need not be made. Evidence is available (Kanal 1974 and Jansen 1981b) concerning the non-stationarity of the EEG indicating that the statistics of the background or ongoing EEG during the time the signal is being generated are different than the statistics of the EEG when no stimulus related signal is being generated. Hence, the estimated covariance matrices will be different and the quadratic classifier should perform better

than the linear classifier.

The covariance was estimated by the unbiased estimator:

$$\hat{\Sigma}_i = \frac{1}{N-1} \sum_{k=1}^{K} (\mathbf{x}_{ki} - \hat{\mathbf{m}}_i)(\mathbf{x}_{ki} - \hat{\mathbf{m}}_i)^T \qquad (2\text{-}3)$$

where $\mathbf{x}_{ki}$ = the kth sample feature vector from class i,

$\hat{\mathbf{m}}_i$ = estimated mean of class i,

and K = quantity of sample feature vectors.

The mean was estimated by the unbiased estimator:

$$\hat{\mathbf{m}}_i = \frac{1}{K} \sum_{k=1}^{K} \mathbf{x}_{ki} \qquad (2\text{-}4)$$

Error analysis was accomplished by computing estimates of the upper and lower error bounds from the data set using the designed classifier. The lower bound of the error was computed by the resubstitution method (C method) where the same data samples that were used to train the classifier (i.e., compute the estimated statistics for the classifier) were also used to test the classifier. The upper error bound was computed by the leaving-one-out method (L method), where the classifier was trained on all but one sample and that one sample was then tested. This was repeated for each sample, each time a different sample is left out of the training set and was then tested (Fukunaga 1972). This was easily implemented for the quadratic classifier by the formulation in (Fukunaga 1971 and 1972).

Feature selection was accomplished by a modified forward sequential feature selection (FSFS) routine (Mucciardi 1971) using the criterion of minimizing the upper or lower bound of the error or a linear combination of the two. FSFS works by first picking the single feature which minimizes the error criterion by testing all features from the set of possible features. Then the next feature is selected by combining each possible feature with the one previously chosen and selecting the best pair. This is further iterated until the maximum number of features is selected. FSFS is not an optimum selection algorithm, but it greatly reduces the computational burden required for an exhaustive search feature selection (ESFS) process. The FSFS routine was implemented in such a manner that alternate combinations of features which produce equally low lower error bounds were tested through the next iteration. This prevented the arbitrary selection of one of the feature combinations while rejecting others which may have been better after the next iteration than the one chosen.

The size of the subset of features selected (N) was determined from the upper error bound. As the number of features selected increases, the upper error bound generally decreases and levels off at some value which was the selected value for N. This error bound generally increases with the selection of additional features (Hughes 1968). The selection of the best set of features was that which had the lowest upper error bound of all the feature sets tested.

The detection problem is set up as a two class classification to distinguish between the two hypotheses H_0 and H_1.

$$\begin{aligned} H_0 &: d(t) = e(t) + n(t), \\ \text{and } H_1 &: d(t) = s(t) + e(t) + n(t). \end{aligned} \tag{2-5}$$

This was implemented by training the classifier for the hypothesis H_1 with the recorded data synchronized with the stimulation. A portion of the EEG data recorded prior to the time of stimulation, the same length as that of the signal, was used to train the classifier for the hypothesis H_0. The section of the EEG used for training the detector for hypothesis H_0 was selected by using a section of the pre-stimulus data whose time relationship to the signal was a random variable to prevent any time-lock effects between the two data sets.

2.3. Data Transformation

Features were selected from a set of the data. The transform action was implemented by sampling the outputs of a bank of bandpass filters. An important part of the design problem was selections of the filter center frequency and shape of the passband.

Referring to Figure 2-2, it is seen that the measured data is first time limited by a window function then transformed into the frequency domain. This results in a short term spectrum computed in the manner described in (Allen 1977). This method has been used extensively in speech analysis systems (Tanaka 1979, Allen 1977, White 1976, Schafer 1973, and Portnoff 1976). The parameters involved are the time window shape and length, and the resolution in the frequency domain. After converting the original data to the frequency domain the 1st classifier/feature selector chooses those frequency components which give the best performance. These results are then used to select the center frequencies of the data transformation filters. The outputs of these filters are used as the [illegible] from which the 2nd classifier/feature selector selects the final feature subset.

The width of the window function was based on the duration of the EP signal which was estimated from the averaged EP. It is desirable to give equal

weighting to all the time points of the sampled signal to prevent distortion of the signal by the window. It is also desirable to include only the data containing the signal in the transformed data so that the transform would reflect only that energy. The window satisfying both these criteria is the rectangular window, but this window produces spurious frequency components in the transform, especially at the higher frequencies. A compromise would not distort the data containing the signal and would allow only a small amount of data not containing the signal to be used in the transformed data. The window which is a combination of a rectangular window and a raised cosine window is a Tukey window (Blackman 1958) of the form

$$w(t) = \begin{cases} \frac{1}{2} + \frac{1}{2}\cos\left[\frac{\pi}{a}(t-t_o)\right], & \text{for } t_o - a \le t < t_o, \\ \frac{1}{2} + \frac{1}{2}\cos\left[\frac{\pi}{a}(t-t_o-b)\right], & \text{for } t_o + b + a > t > t_o + b, \\ 1, & \text{for } t_o \le t \le t_o + b, \\ 0, & \text{otherwise.} \end{cases} \tag{2-6}$$

The parameter a defines the width of the raised cosine segments at the ends of the constant segment. The parameter b defines the length of the constant segment and t_o defines the beginning of the constant segment. The value chosen for parameter a was 80 ms. This was a compromise between large values which allow non-signal data to be used in the transform and small values which produce the spurious frequencies due to the edge effect of the window. The constant segment parameter b was set to the length of the signal found in the average EP. An example is displayed in Figure 2-3 in which an averaged EP from the visual stimulus experiment is plotted along with the time window selected.

Desirable characteristics for frequency features used for classification be that they are Gaussian distributed and that they represent the information at that frequency efficiently for class separation. Features which are the real or imaginary parts of the FFT transformed data were found to have approximately Gaussian amplitude distributions (Moser 1984). Two features represented the information at each frequency. They were obtained directly from the FFT and the frequency transform block in the block diagram of the 2-step system (Figure 2-2) was bypassed. Examples of the distribution of the frequency data are shown in Figure 2-4. Figure 2-4a portrays 100 points from each of 2 classes at 2 Hz and Figure 2-4b is a similar plot at 4 Hz. Evident in

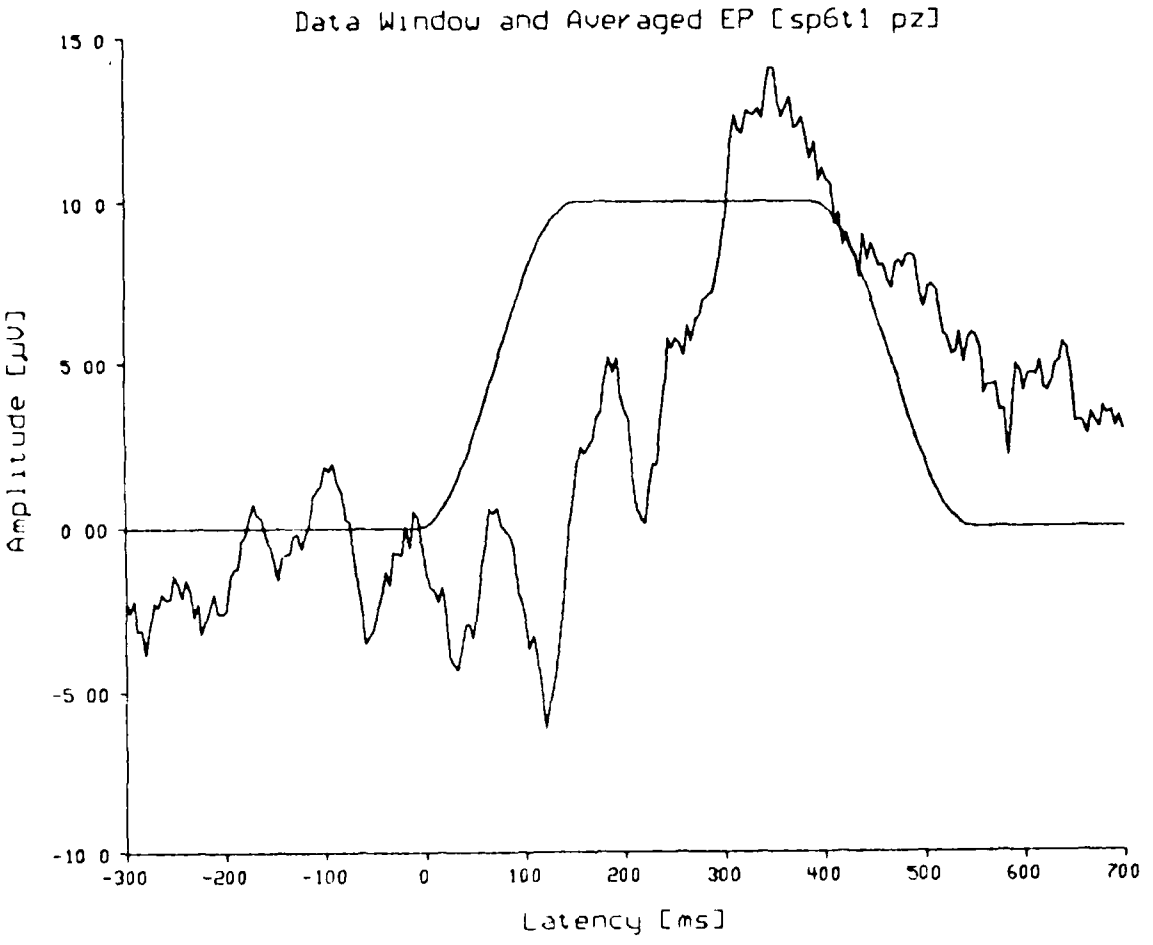


Figure 2-3 A Typical Averaged Evoked Potential Signal from the Sternberg Paradigm Experiment Plotted Along With the Tukey Window Applied to This Type of Data

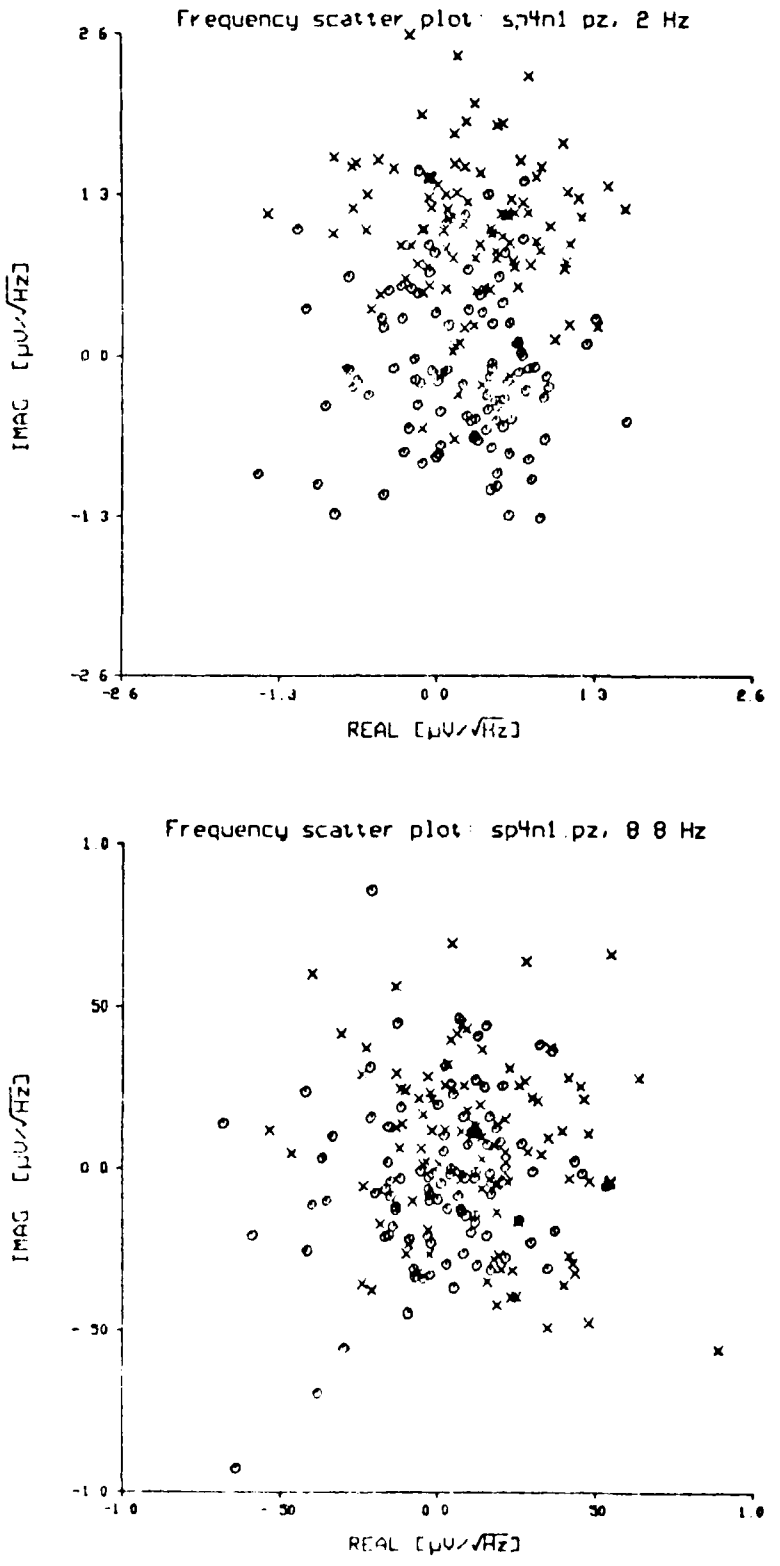


Figure 2-4 Scatter Plot Examples of the Frequency Domain Features for Two Classes of Signals

this figure, as was typical for the frequencies from the various data sets, is the inherent lack of separability for single frequencies. There is higher separability of the classes when higher dimensional distributions are considered. Dimensionality of up to 10 features was considered for most of the feature selection processing. Since the real and imaginary parts are each features, there were twice as many features as frequencies.

A procedure was developed to transform features from the FFT by projecting the complex frequency data onto a line. The new features X(f) at various frequencies f are computed by:

$$X(f)=M(f)\cdot\cos[\phi(f)-\theta(f)] \tag{2-7}$$

where $M(f)$ = Magnitude at frequency f,

$\phi(f)$ = Phase angle at frequency f,

and $\theta(f)$ = Reference angle at frequency f.

The reference angle was determined from the line which connects the means of the 2 classes in the complex frequency plane. This maximizes the mean separation between the classes.

The frequencies of the features producing the best results with the first classifier/feature selector were used to select the center frequencies of the filters which process the raw data for use by the second classifier/feature selector. Non-causal, symmetric, finite impulse response bandpass filters were used to filter the raw data in the second step of the 2-step process. This type of filter has a zero phase response and renders only a small portion of the ends of the data, equal to one half the filter length, unreliable after filtering. The filter h(t) was designed in the time domain by multiplying a sinusoid of the appropriate frequency and phase by a time window:

$$h(t) = w(t + t_o) \cdot \sin(2\pi f_o t + \phi) \tag{2-8}$$

where t_o = time location of window,

f_o = sinusoidal frequency,

and ϕ = sinusoidal phase.

This is equivalent in the frequency domain to convolving a pair of impulses at frequency f_o and $-f_o$ with the transform of the window function:

$$H(f) = W(f) * \frac{1}{2}\left[\delta(f - f_o) + \delta(f + f_o)\right] \qquad (2\text{-}9)$$

where $H(f)$ = Fourier transform of $h(t)$,
and $W(f)$ = Fourier transform of $w(t)$.

The outputs from the filters were used as a large feature set where the amplitudes at the times points of each filter were the features (see Figure 2-3). The number of filters was set by the number of different frequencies selected by the first classifier/feature selector. The chosen features were amplitudes of selected filters outputs at selected times.

2.4. Data and Testing

The procedures for implementing the proposed processing techniques were tested with artificially generated data and actual human data from EP experiments. The human data experiments are discussed in this section and the results using various artificial data sets are discussed in Appendix A.

The Sternberg paradigm was selected for generating experimental EP waveforms used in this research (Sternberg 1966). The details of the Sternberg experiment were summarized in a previous technical report (McGillem 1983). The plots of the averages of the EP's for electrode Pz for the four subjects used in subsequent processing are presented in Figure 2-5.

The data from the 4 subjects were processed by the 2-step process with fixed filter bandwidths of 5, 10, and 15 Hz, and proportional bandwidths of 0.25, 0.50, 0.75, and 1.00 times the center frequency of the filter. The bandwidths are measured from null to null in the magnitude spectrum, which gives a larger value than when measured between the -3dB points of the spectrum. Tables 2-1 and 2-2 list the highest lower accuracy bound estimates along with the number of features selected to achieve this accuracy for the raw data, frequency data (step 1 of the 2-step process) and filtered data (step 2). The averages of the results across the 4 subjects are listed in the last column of these tables. Table 2-1 lists the results of processing the frequency data without transformations, which allows the real and imaginary part of the frequency data to be used directly by the first classifier/feature selector. In this case, there is twice the number of frequency features from which to choose. Table 2-2 lists the results of processing using the frequency transformation that projects the frequency data onto a line in the complex frequency plane, yielding 1 feature per frequency.

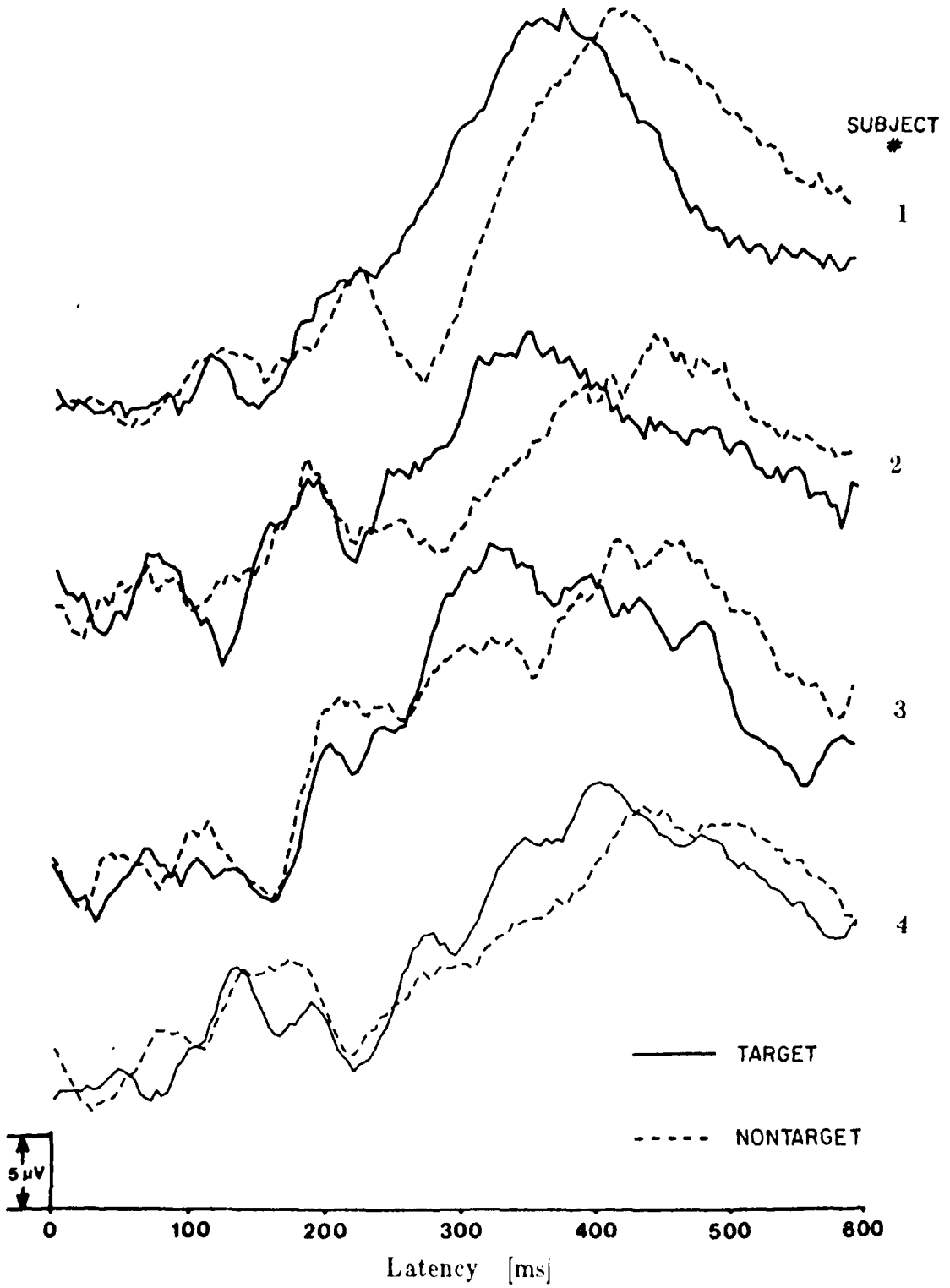


Figure 2-5 Averaged Evoked Potentials for Electrode Pz, 4 Subjects, Target vs. Non-Target, 1 Target in Target Set

Table 2-1 Results From the 2-Step Classifier/Feature Selection Process Without Frequency Feature Transformation. Table Entries: Percent Classification Accuracy / Number of Features.

		Subject				
Data	BW	1	2	3	4	Average
Target	Raw	93.5/9.5	77.9/4	88.8/7	88.3/5	87.1/6.5
	Frequency	96.5/8	79.6/4	94/8	90.3/6	90.1/6
	5 Hz	95/5	79.1/2	94.5/9	91.8/6	90.1/5
	10 Hz	97.5/6	79.6/4	94.5/10	98.5/7	92.5/7
	15 Hz	98/3	76.5/2	96/9	99/7	92.4/5
	X.25 Fo	95/6	74/1	90.5/4	88.3/2	87.0/4
	X.50 Fo	93.5/4	78.1/2	99/8	91.3/3	90.5/4.5
	X.75 Fo	94/7	80.1/8	97.5/10	92.9/8	91.1/8
	X1.0 Fo	98.5/10	86.7/5	99.5/8	89.3/3	93.5/6
	Best	X1.0	X1.0	X1.0	15 Hz	X1.0
Nontarget	Raw	84.5/8	76/5	92/8	80/4	83.1/6
	Frequency	97/7	89/6	94.5/6	89/5	92.4/6
	5 Hz	95/5	81.5/7	93/3	92/4	90.4/5
	10 Hz	98.5/6	86/7	99.5/6	92.5/10	94.1/7
	15 Hz	96.5/3	97/7	97.5/7	100/8	97.8/6
	X.25 Fo	87.5/5	86.5/4	96.5/6	93/8	90.9/6
	X.50 Fo	99/8	82/3	96.5/5	90/4	91.9/5
	X.75 Fo	98.5/9	84.5/10	98/8	91/10	93.0/8.5
	X1.0 Fo	99/5	80.5/3	99.5/5	95.5/10	93.6/5.5
	Best	X1.0	15 Hz	X1.0	15 Hz	15 Hz
Target vs. Nontarget	Raw	73.5/4	70.4/8	73.5/7	69.9/5	71.8/6
	Frequency	77/6	74.5/10	75/4	73/4	74.9/6
	5 Hz	75.5/2	78.6/10	74.5/4	70.9/2	74.9/4.5
	10 Hz	78/4	78.6/10	76.5/4	71.9/3	76.3/5
	15 Hz	72.5/4	76/5	72/5	66.3/7	71.7/5
	X.25 Fo	75/2	76/10	74.5/2	71.9/5	74.4/5
	X.50 Fo	73.5/2	80.6/6	75/3	67.9/2	74.3/3
	X.75 Fo	74.5/9	74.5/7	76/4	67.9/2	73.7/5.5
	X1.0 Fo	75.5/4	79.1/5	78.5/7	66.8/2	75.0/4.5
	Best	Freq	X.50	X1.0	Freq	10 Hz

Table 2-2 Results From the 2-Step Classifier/Feature Selection Process With Frequency Feature Transformation. Table Entries: Percent Classification Accuracy / Number of Features.

		Subject				
Data	BW	1	2	3	4	Average
Target	Raw	93.5/9.5	77.9/4	88.8/7	88.3/5	87.1/6.5
	Frequency	98.5/9	84.2/8	94/9	90.3/5	91.8/8
	5 Hz	93/3	83.7/9	92/8	90.3/7	89.8/7
	10 Hz	99/10	95.9/9	99.5/7	100/8	98.6/8.5
	15 Hz	95.5/6	94.9/10	100/7	100/5	97.6/7
	X.25 Fo	95.5/9	80.1/5	89.5/3	90.8/3	89/5
	X.50 Fo	99.5/8	99.5/10	99/10	93.4/8	97.9/9
	X.75 Fo	97/9	99.5/8	98.5/9	92.4/10	96.9/9
	X1.0 Fo	99/6	96.4/5	99.5/8	98.5/9	98.4/7
	Best	X0.50	X0.75	15 Hz	15 Hz	10 Hz
Nontarget	Raw	84.5/8	76/5	92/8	80/4	83.1/6
	Frequency	90.5/9	80/3	95/6	87.5/7	88.3/6
	5 Hz	88.5/3	84.5/3	97/8	85/5	88.8/5
	10 Hz	88.5/3	97.5/10	100/6	89.5/5	93.9/6
	15 Hz	99/8	98.5/10	100/6	92/10	97.4/8.5
	X.25 Fo	87.5/7	84.5/4	95.5/6	79/6	86.6/6
	X.50 Fo	86/4	82/5	97.5/8	94.5/10	90/7
	X.75 Fo	91.5/10	83.5/2	99/7	90/9	93.5/7
	X1.0 Fo	97.5/10	83/4	100/6	99/9	94.9/7
	Best	15 Hz	15 Hz	Several	X1.0	15 Hz
Target vs. Nontarget	Raw	73.5/4	70.4/8	73.5/7	69.9/5	71.8/6
	Frequency	78/3	76.5/9	73.5/6	75.5/4	75.9/5.5
	5 Hz	79/6	80.6/3	77/3	74/4	77.7/4
	10 Hz	76/2	79.6/5	75.5/2	71.9/4	75.8/3
	15 Hz	78/9	77/7	73.5/5	71.9/6	75.1/7
	X.25 Fo	78/7	80.1/6	78.5/4	74/5	77.7/5.5
	X.50 Fo	77/3	82.7/9	77/4	69.9/3	76.7/5
	X.75 Fo	77.5/5	82.1/6	77/3	72.5/6	77.3/5
	X1.0 Fo	80/9	80.1/8	75.5/7	70.9/4	76.6/7
	Best	X1.0	X0.50	X0.25	Freq	5 Hz

In both cases the data window in the first step of the process extended from 0 ms to 556 ms latency with the flat portion extending from 160 ms to 396 ms, which was also the range over which filtered amplitude features were searched in the second step of the 2-step process. Sixty frequency features ranging from 0.98 to 58.6 Hz were used. The frequency spacing was approximately 1 Hz. In most cases, 100 records from each class were used. Ninety-eight records were used in the cases where a full set of 100 good records was not obtained.

Each table includes the results for detection of the target set EP data (t), detection of the non-target set EP data (n), and classification between the target and non-target sets (t/n). For detection, the ongoing EEG class (no EP) was taken as the section of data starting at 760 ms before stimulation. This starting point was randomly varied across the data records over a uniformly distributed range of ±100 ms.

The best averaged accuracy results for detection of the targets are as follows:

1) 98.6%, BW=10 Hz, with frequency transformation,
2) 98.4%, BW=X1.0, with frequency transformation,
3) 97.9%, BW=X0.5, with frequency transformation,
4) 97.6%, BW=15 Hz, with frequency transformation.

The bandwidths with a leading X are proportional bandwidths, the number indicating the proportion of the center frequency, with a minimum bandwidth 4 Hz. These results show a significant improvement over the use of raw amplitude features which produced an averaged accuracy of 87.1% across the subjects. The best averaged accuracy result for frequency features was 91.8% using frequency feature transformation, 3.7% higher than for raw amplitude features. For some subjects lower bound accuracies were 100%. As much as 20% or more improvement was achieved for the filtered data features over raw data features in some cases, as in subject 2, BW=X0.5 or X0.75, with frequency transformation (Table 2-2), where 21.6% improvement was achieved. If these results were presented in terms of error instead of accuracy, this case would improve the error from 22.1% to 0.5%, a significant improvement.

The best averaged accuracy results for detection of the non-target EP's are as follows:

1) 97.8%, BW=15 Hz, without frequency transformation,
2) 97.4%, BW=15 Hz, with frequency transformation,
3) 94.9%, BW=X1.0, with frequency transformation,
4) 94.1%, BW=10 Hz, without frequency transformation.

These results also show a significant improvement over the averaged accuracy of 83.1% for raw amplitude features. For frequency features, the best averaged accuracy was 92.4%, without frequency transformation, a 9.3% improvement over raw amplitude features. Improvements in the accuracy for using filtered features over unfiltered features were over 20% in some cases as in subject 2, BW=15 Hz, with frequency feature transformation, in which case the error went from 24% to 1.5%. In all cases for detection, the accuracy achieved using raw amplitude features was improved by using filtered features determined by the 2-step classifier/feature selection process.

The best averaged accuracy results for classification between the two classes are as follows:

1) 77.7%, BW=5 Hz, with frequency transformation,
2) 77.7%, BW=X0.25, with frequency transformation,
3) 77.3%, BW=X0.75, with frequency transformation,
4) 76.7%, BW=X0.50, with frequency transformation.

These results show a small improvement over the unfiltered data features averaged accuracy of 71.8% of up to 5.9%. The best averaged accuracy for frequency features is 75.9%, with frequency transformation, a 4.1% improvement over those achieved using raw amplitude features. The best filtered data averaged result is 1.8% better than this. Improvements of up to 11.7% were achieved for individual subjects as in Subject 2, BW=X0.50, with frequency feature transformation.

From this study on a limited number of 4 subjects and data sets, it is evident that the 2-step method produced very high accuracies for detection, representing large improvements over 1-step methods using unfiltered or lowpass filtered amplitude features. Since the estimated detection accuracies were as high as 100% or very near this level, the method is an excellent way to determine whether single EEG records contain a particular EP. This may be used as the basis for an on-line detector. The classification results showed smaller gains with the use of filtered over unfiltered amplitude features. These could possibly be improved further by testing a wider range and finer resolution of parameter values such as bandwidths, data window parameters, and number of frequency features. Furthermore, different filter responses could be tested including non-symmetric ones. The preliminary results look promising considering the limited number of parameter values tested.

As described in Appendix A, four different types of signals were used in artificial data sets to test the 2-step classifier/feature selector and compare the results to those using 1-step techniques (raw features or frequency features).

From a limited study on these signals in actual EEG data at different SNR levels, it was found that the filtered features of the 2-step method produced significantly improved accuracies (see Table A-5 for a summary of the results). For detection at the 0dB level, raw features produced very high accuracies of 96.1% to 99.5%. The best results using frequency features yielded accuracies of 99%-100%, and 100% accuracies were achieved for all signals using filtered features. For detection with SNR's of -6dB, raw features produced accuracies ranging from 78.5% to 92.5%. The best results using frequency feature yielded accuracies ranging from 81% to 93.5%. This represents small improvements over using raw features. The best results using filtered features showed significant improvements in these accuracies, yielding 100% for 4 of the 8 data sets tested. This represents improvements up to 19.5%, probably limited by the fact that the maximum accuracies of 100% were achieved in many cases. These improvements are also much larger than those obtained using frequency features.

For detection of the signals in EEG at SNR's of -12dB, raw features produced average accuracies ranging from 66.2% to 83.5%. The best results using frequency features showed improvements to these accuracies of -1.9% to 6.8%. Filtered features again showed substantial improvements of up to 31.8%, probably limited by the fact that near maximum accuracies were achieved in all cases.

Classification between the artificial data sets also showed similar improvements by using the 2-step procedure. At SNR levels of 0dB, the best frequency features results yielded 0.5% to 10.2% improvements, and the best filtered features results yielded 0% to 10.2% improvements over the results using raw features. These improvements were probably limited because the accuracies achieved were at or near 100%. At SNR levels of -6dB, the best frequency features yielded improvements of 3.1% to 12.2%, and the best filtered features yielded 6.6% to 13.8% improvements over the results obtained using raw features. Similar improvements were achieved at SNR levels of -12dB with 1.6% to 10.2% improvements afforded by the best frequency features results, and 5.6% to 13.2% improvements afforded by the best filtered features results.

The 2-step process has consistently afforded very high detection accuracies in the data sets tested, even in EEG noise at SNR's as low as -12dB. It is stressed that the accuracies are estimates of the lower bound of the Bayes accuracy (upper bound of the error), and it is expected that the actual Bayes accuracy should not be lower than these figures. The consistency and

magnitude of the improvements is strong evidence of the validity and applicability of the technique, especially for low SNR levels where the signal power is smaller than the noise power.

Comparing the results from testing human data to those from testing artificial data, the degree of improvement afforded by using filtered features over raw features for detection were similar, especially for the -6dB SNR artificial data. Improvements on the order of 20% were obtained in both cases. For both the human and artificial data, frequency features generally produced improvements over raw features, but not to the extent of the filtered features. The accuracy levels achieved for the human data for the various features falls somewhere between those of the artificial data at SNR's of -6dB and -12dB.

The number of features used to achieve the resulting accuracies were similar for the various types of features. The possibility that the improved accuracies for frequency or filtered features results from the use of a larger combination of features was considered. There is a large range in the number of features used, but there is no consistent trend of a much larger set of filtered features being necessary to achieve the performance increases. In almost all cases, the maximum lower bound accuracy was achieved with less than the maximum of 10 features selected for both the human and artificial data. Therefore, the maximum number of 10 features was indeed sufficiently large to prevent this parameter from being a limiting factor in the maximum achievable accuracy.

When accuracies were very high for artificial data with SNR's of 0dB, only small improvements could be attained with filtered features. But the number of filtered features necessary to achieve the high accuracies was lower than the number of raw features, in some cases 100% was achieved with only 1 or 2 filtered features.

It is not clear which is the single best bandwidth to use for the filtered features. There are some indications, but with no obvious trend, that the larger bandwidths produce better accuracies of detection for both the human and artificial data. The lowest bandwidth of 5 Hz and X0.25 did the poorest. The distinction amongst the others is not clear. The number of different bandwidths tested was limited due to the large amount of computing time required for each test. A fairly large range was covered by the choice of bandwidths, but it may not have been wide enough or of fine enough resolution to indicate the best bandwidths. It may be that even higher accuracies can be achieved by using different bandwidths for the different filters in step-2, but this presents the problem of how these bandwidths would be chosen. This

would extend the 2-step method to a 3-step method, where the chosen parameters are frequency and time, plus the parameter of bandwidth. The computational time complexity of this type method would be much larger than the 2-step method, limiting its usefulness.

2.5 Conclusions

This work focused on the development of new methods for extracting information from evoked potential data using time-frequency amplitude features. The methods developed show promise for improving the detection and classification abilities of the computer in distinguishing single EP's over conventionally used methods. Classification has generally been improved approximately 6% over using raw amplitude features for the limited amount of data which was tested

Results for detection were very encouraging, with improvements over using unfiltered data on the order of 20%. This improvement may have been influenced heavily by the "phase-alignment" properties of the EP as presented in (Jervis 1983). In this work, the EP and EEG data were transformed into the frequency domain and a particular frequency was plotted on the complex plane. For the EEG data, the points corresponding to the various sample records had phases which were fairly evenly distributed (i.e., clustered about the origin). The EP data points had phases which were distributed more closely about a mean phase, resulting in a plot with the points clustered about a point offset from the origin. The 2-step method capitalizes on this separation of the 2 classes in the frequency domain by performing classification based on frequency information, and by performing a frequency feature transform which attempts to maximize the separation of the classes in the frequency domain.

Alteration of parameter values and filters may increase this improvement further. Although the actual detection or classification is almost instantaneous, the analysis leading up to the design requires a substantial amount of computer time due to the feature selection process. Therefore only a limited variation of parameters was investigated. At this time there is no direct method for determining optimum values. Further research could concentrate on methods to determine improved filter functions which could yield more precise information. The bandwidths of the filters in this study were either fixed or had a fixed proportionality constant for the various center frequencies. Improved results may be obtained by using a combination of different bandwidths, possibly determined by combining the best features from tests

using different bandwidths.

Improvements to these methods may result from improvements in amplifier design. Many of the frequency features generally selected had frequencies above 25 Hz. At these frequencies, the SNR becomes low since the spectrum of the EEG falls off rapidly with higher frequencies while the amplifier noise is essentially constant. The quieter the amplifiers are, the more effectively the information in the higher frequencies can be used.

3. OPTIMAL AND SUBOPTIMAL FEATURE SELECTION FOR QUADRATIC CLASSIFICATION

3.1. Introduction

This preliminary research investigated the exhaustive search feature selection (ESFS) procedure using a quadratic discriminant function for classification. The quadratic Bayes classifier under the Gaussian assumption (as described in the previous chapter) was used to evaluate the feature selection procedures. A previous investigation compared the ESFS performance with conventional feature selection procedures using a linear discriminant function (McGillem 1983 and Halliday 1985). In that research, an efficient algorithm for implementing the ESFS error criterion was derived which greatly reduced the computational burden of the ESFS algorithm. The quadratic discriminant function does not lend itself to efficient evaluation; therefore, no attempt was made to derive an efficient algorithm. The ESFS feature selection was accomplished by examining all possible combinations of features at each level of selection (i.e., 1 feature, 2 features, 3 features, etc.).

3.2. Methods

Error analysis was accomplished by computing upper and lower error bounds from the data set using the designed classifier. The lower bound of the error was computed by the resubstitution method (C method) where the same data samples that were used to train the classifier (i.e., compute the estimated statistics for the classifier) were also used to test the classifier. The upper error bound was computed by the leaving-one-out method (L method), where the classifier was trained on all but one sample and that one sample was then tested. This was repeated for each sample, each time a different sample is left out of the training set and was then tested (Fukunaga 1972). This was easily implemented for the quadratic classifier by the formulation in (Fukunaga 1971 and 1972).

Feature selection was accomplished by a modified forward sequential feature selection (FSFS) routine (Mucciardi 1971) using the criterion of minimizing the upper or lower bound of the error or a linear combination of

the two. FSFS works by first picking the single feature which minimizes the error criterion by testing all features from the set of possible features. Then the next feature is selected by combining each possible feature with the one previously chosen and selecting the best pair. This is further iterated until the maximum number of features is selected. FSFS is not an optimum selection algorithm, but it greatly reduces the computational burden over using an exhaustive selection process. The time complexity of FSFS is approximately $0(n \cdot m^3)$ where m is the number of features to be selected and n is the number of possible features from which to select. The computational burden for an exhaustive search is $0(n^m \cdot m^3)$. This results in a large difference in computational burden for selecting sets of up to 10 features as was done in this study. The term m^3 in the time complexity figures is due to the mxm matrix inversion. The time complexity for FSFS is somewhat more complex than shown but this figure is a good approximation for m greater than 3 or 4. The other terms would be small compared to the one shown because they would be of the form $n \cdot (m-k)^3$ where $1 \leq k \leq m-1$.

The FSFS routine was implemented in such a manner that alternate combinations of features which produce equally low lower error bounds were tested through the next iteration. This prevented the arbitrary selection of one of the feature combinations while rejecting others which may have been better after the next iteration than the one chosen .

Another option implemented in the feature selection algorithm is to allow the choosing of the 1st k features via an exhaustive search and then continuing to pick the rest of the features by FSFS. The computational burden of this method would be approximately $0(n^k k^3 + nm^3)$. This idea has been previously suggested (Lissack 1976) for an exhaustive search for the best set of 2 features. This option allowed testing various tradeoffs between decreased error and greater time complexity for increasing k. Further comparisons could be made between the sub-optimal feature selection process and the optimal exhaustive feature selection process implemented by setting k equal to the maximum number of features to be selected. The exhaustive search and other feature selections with various k values were run on selected data sets to observe the characteristics of the tradeoffs and to compare the results of the sub-optimal search procedures to the optimal one. The testing provided a good indication of how well the sub-optimal methods perform.

The size of the subset of features selected (m) was determined from the upper error bound. As the number of features selected increases, the upper error bound generally decreases and levels off at some value which was the

selected value for m. This error bound then increases with the selection of additional features (Hughes 1968). The selection of the best set of features was that which had the lowest upper error bound of all the feature sets tested.

3.3. Results and Conclusions

The following questions were examined while comparing the feature selection procedures:

1) whether basing feature selection on the computed upper accuracy bound (C) produces as high classification accuracies as basing feature selection on the lower accuracy bound (L), or on a combination of the two bounds,
2) whether using forward sequential feature selection (FSFS) significantly reduces the resulting classifier accuracies over using exhaustive search feature selection (ESFS),
3) whether some combination of ESFS and FSFS and basing the feature selection on criteria other than just the upper bound would significantly improve the classification accuracies.

The computational burden for the computation of the lower accuracy bound is very small and the use of this bound alone or in combination with the upper bound would be justified for small gains in the resulting accuracies. The computational burden of ESFS is extremely high for selecting a moderate number of features as described in the introduction. However, there may be advantages in selecting a small number of features by ESFS and then selecting additional ones by FSFS.

To reduce the number of possible amplitude features from which to select, the raw data were low-pass filtered to 25Hz (20Hz cutoff) and resampled at 50 samples per second. This process yielded one fifth the number of features as the original data which was sampled at 250 samples, allowing the ESFS to be completed in a reasonable amount of time. The data used was from a pilot run of the Sternberg paradigm experiment. Only the condition of 1 target in the target set was run in this experiment. The data were contaminated by a small amount of 60Hz interference from the video monitor which was inside the testing chamber, but this was not a problem since the data were lowpass filtered to 55Hz, and then again to 25Hz.

Data from electrodes O_z and C_z were used in the classification test, and a combined data set of data from electrodes O_z, P_z, and C_z was also tested. Fifteen amplitude features ranging from 160 ms to 460 ms were employed and the results tabulated in Table 3-1. The leftmost column lists the number of

Table 3-1 Results From Subject #1 Amplitude Data for Various Levels of Exhaustive Search Feature Selection and Selection Criteria

Table entries: upper bound accuracy (percent)/lower bound accuracy (percent)/number of features.

No. of features selected by ESFS	Feature Selection Criteria Weight		Amplitude features - lowpass filtered and undersampled			Amplitude features - unfiltered	
	C	L	Oz	Cz	Combined Oz,Pz,Cz	Oz	Pz
0	1	0	76/73/5	73/68/7	87/77/10	84/81/8	84/76/8
	1	1	76/73/5	72/68/8	83/78/10	86/81/10	78/76/7
	0	1	76/74/6	73/69/8	81/78/9	84/80/10	80/77/8
2	1	0	79/73/5	73/68/7	78/74/9	82/75/8	83/79/8
	1	1	76/73/5	72/68/8	79/74/10	83/78/10	82/79/8
	0	1	76/73/5	73/69/8	81/78/9	83/80/10	77/75/6
3	1	0	76/73/5	77/67/9	83/77/10	83/81/9	82/74/8
	1	1	76/73/5	73/69/9	83/78/10	83/81/9	82/75/8
	0	1	76/74/6	74/67/9	81/78/9	80/79/7	80/75/7
4	1	0	76/73/5	75/69/8			
	1	1	76/73/5	73/69/9			
	0	1	76/73/5	76/69/10			
5	1	0	76/73/5	85/81/8			
	1	1	76/73/5	73/73/5			
	0	1	76/73/5	76/74/6			

features first selected by ESFS, the rest of the features (up to a total of 10) were selected by FSFS. The first row of entries is for zero features selected by ESFS and all selected by FSFS. This value may have also been labeled 1 because FSFS and ESFS are equivalent for the selection of 1 feature. The next 2 columns are the relative weightings given to the upper (C) and lower (L) accuracy bounds. The table entries are the upper bound accuracy followed by the lower bound accuracy followed by the number of features which produced the listed highest lower bound accuracy. The righthand columns of entries list the accuracies for raw amplitude features, with features selected from a set of 75 features. Results were not computed for ESFS over 3 features for the data sets from combined electrodes data and raw data sets because the computation time became very great with these larger numbers of features from which to select for the larger ESFS values.

The general conclusion drawn from these results is that there is no advantage to using the lower accuracy bound for feature selection, or in performing ESFS. For the filtered data, electrode O_z produced almost uniform results across the various selection conditions with lower bounds of 73-74%. The features which were selected were the same in most cases. These results are similar to those of electrode C_z, although this data set produced more variability across the conditions. The lower bound accuracy varied from 67 to 81%, mostly restricted to 67-69%. Five features selected exhaustively produced the best accuracy bounds of 85% and 81% for selection based on the upper bound. This was the only example where ESFS produced significantly better results (but at a high computational cost). When the lowpass filtered and undersampled amplitude data from 3 electrodes were used as features from which to select, bound accuracies ranged from 74% to 78% for the lower bound and 79% to 87% for the upper bound. In all these cases, the results for the first row of ESFS=0, C=1, L=0, indicate that this is as good a choice as almost any other of the conditions for feature selection. Hence, FSFS is used as the feature selection method by the 2-step procedure.

The above analysis of the results on the filtered data also applies to those of the raw amplitude data. The lower bound accuracies ranged from 75% to 81%, and the first row results are at the upper end of the range. These results for the unfiltered data are significantly better than the corresponding results for the filtered data, ranging from 2% to 8%, mainly 5-8%. More features were chosen to achieve these higher accuracies, but there were more from which to choose. Thus, heavy filtering may diminish pertinent information contained in higher frequencies and, therefore, be detrimental to classification accuracy.

This is an important reason for including frequencies beyond 25Hz in the 2-step method. Adding more features would further reduce the ESFS advantages because the additional features would lead to a substantial increase in the time required to exhaustively search for the optimum feature set.

4. MULTICHANNEL ERP WAVEFORM ESTIMATION

4.1. Introduction

Research into the theory and design of event related potential (ERP) estimation filters has been extended to include the information available in multiple electrode recordings. The theoretical work for the continuous case optimum linear time-varying estimation filter was developed by Booton (1952) and extended to multiple channels by Wolf (1959). The single channel discrete ERP filter was derived by Yu (1982 and 1983) and described in an earlier technical report (McGillem 1983). A detailed analysis of the results of the discrete multichannel time-varying filter (MTVF) research is available in the doctoral dissertation of Westerkamp (1985). This chapter summarizes that research. The theoretical basis of the procedure is reviewed first, a description of the implementation is presented, and finally experimental results using both simulated and human ERP data are presented and discussed.

4.2. Theoretical Analysis

The optimum linear MTVF can be formulated by requiring that the signal estimate be a linear transformation of the input received data.

$$\underline{\hat{s}}_d = \underline{H}\underline{r} \tag{4-1}$$

where $\underline{\hat{s}}_d$ = the estimate of the desired signal,

$\underline{H}$ = the filter matrix,

$\underline{r}$ = the input multichannel data.

The desired signal is typically the signal present in one of the input channels. Figure 4-1 depicts a block diagram of the MTVF. Although it is not necessary to assume that the input is an additive signal plus noise process, this assumption is made to derive a more powerful filter. The assumptions used to derive the optimum filter are summarized subsequently. The input at each

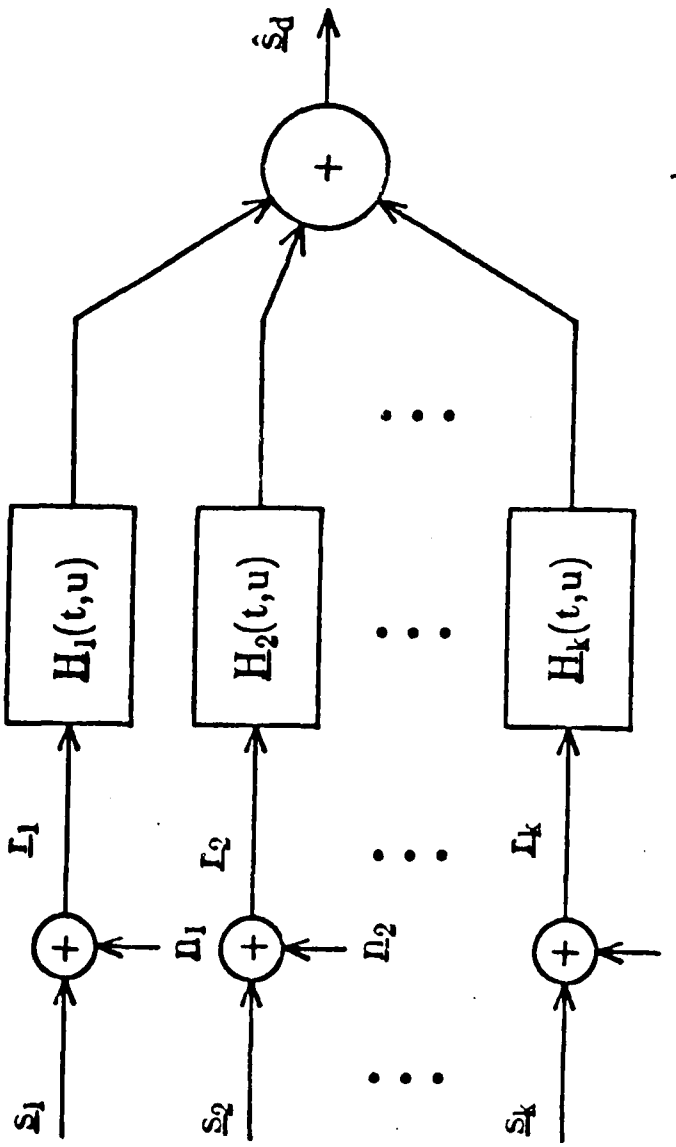


Figure 4-1 Block Diagram of the Multichannel Optimum Linear Filter

channel is passed through a time-varying filter matrix and the output of each filter is summed to provide the desired signal estimate. It is implicit that each filter matrix uses information from all appropriate channels to determine its output.

The optimum MTVF is derived by minimizing the mean-square error between the output of the filter and the desired signal. By including *a priori* information in the minimization process, a more powerful filter can be derived. The following assumptions concerning the EP (signal) and EEG (noise) are made:

(1) the signal and noise processes in each channel are additive,

(2) the signal processes are uncorrelated with the noise processes in and across each channel, and

(3) the noise processes are zero mean.

Combining assumptions (2) and (3), it can be concluded that the signal and noise processes are orthogonal. All cross-correlation matrices between a signal process and a noise process will therefore be zero and drop out of the derivation. The resulting set of simultaneous linear equations which must be solved to obtain the optimum MTVF can be grouped into one large matrix equation as

$$\underline{H}\underline{R}_{rr} = \underline{R}_{s_d s} \tag{4-2}$$

where $\underline{H}$ is a block row vector containing the channel filter matrices, $\underline{R}_{rr}$ is a block matrix containing the cross-correlation matrices among the input data channels, and $\underline{R}_{s_d s}$ is a block row vector containing the cross-correlation matrices between the desired signal and the signals in each of the input channels. If there are k channels and N data points per input data record, $\underline{R}_{rr}$ will be a k×k block matrix. Each block of $\underline{R}_{rr}$ will be an N×N cross-correlation matrix between two of the input channels. Similarly, $\underline{H}$ is a 1×k row vector of N×N filter matrices, and $\underline{R}_{s_d s}$ is a 1×k block row vector of N×N cross-correlation matrices between the desired signal and the signals in each of the input channels. Note that $\underline{R}_{rr}$ is a symmetric matrix while the others in

Equation (4-2) are not (since they are not square matrices).

Theoretically solving Equation (4-2) for $\underline{H}$ gives

$$\underline{H} = \underline{R}_{s_d s} \underline{R}_{rr}^{-1} \quad . \tag{4-3}$$

It is very inefficient to solve a set of simultaneous linear equations by inverting the coefficient matrix. Gaussian elimination (using the symmetry of $\underline{R}_{rr}$) is the method of choice if the equations are well-behaved. Methods such as Levinson recursion which take advantage of Toeplitz coefficient matrices are not applicable here because the input processes may be nonstationary. Experience has shown, however, that the system of equations is not well-behaved and that simplistic solutions often lead to sub-optimum filters which can become unstable. These problems have stimulated research into stable implementations of the optimum filter involving unique pseudoinverse solutions to the filter matrix equation. One such implementation is discussed in the next section.

4.3. Implementation

The optimum filter solution to the matrix equation is often undefined or leads to a filter whose output can become unstable. This is because the matrix $\underline{R}_{rr}$ can be ill-conditioned with respect to inversion. Since matrices may be ill-conditioned with respect to many different algebraic operations, it is assumed throughout the rest of this report that ill-conditioned means with respect to inversion. The ill-conditioned matrix is the result of (1) the ill-posed nature of the original optimum filter problem and (2) the limited precision of the computer operations. These two problems are considered separately.

An ill-posed problem is one whose solution is highly unstable and extremely sensitive to small changes in the design parameters. In the continuous case, the optimum linear filter is the solution to a linear Fredholm integral equation of the first kind. This integral equation is a well-known ill-posed problem. Van Trees (1968) states that, in the absence of white noise, a bounded solution to the integral equation does not exist. Hanson (1971) has derived numerical methods for solving Fredholm integral equations of the first kind and Varah (1973) has described solutions to the general ill-posed problem. In the case of the *a posteriori* optimum MTVF, the ill-posed nature of the problem is the result of three conditions: (1) the absence of a white noise component, (2) the use of estimates rather than exact an knowledge of the

input statistics, and (3) errors in the assumptions used to derive the filter. The effect of the first condition was considered above. The second condition leads to a solution which is only an estimate of the optimum MTVF. Since the solution is only an estimate, it has bias and variance problems which degrade the filter's performance. The third condition refers to the assumption that

$$\underline{R}_{rr} = \underline{R}_{ss} + \underline{R}_{nn} \tag{4-4}$$

It is important that the $\underline{R}_{rr}$ in Equation (4-4) be used when designing the filter. Using the input data records to estimate $\underline{R}_{rr}$ is incorrect because, if the assumptions used to derive the filter are wrong, the designed filter will not be optimum and its output might not minimize the mean-square error criterion. It is better to design the filter using Equation (4-4) and then optimize the design to account for the differences between the assumed input processes and the actual input. This is the basis of the implementation described subsequently.

A problem which is not ill-posed can still result in a coefficient matrix which is ill-conditioned. The computed inverse may be significantly different from the inverse of the original matrix due to roundoff error and finite machine precision. Even though the original matrix is invertible, therefore, it may be ill-conditioned as far as the computer is concerned (Wilkinson 1963). According to Strang (1976), this undesirable result occurs because linear algebraic theory assumes that the matrix operations are performed on a closed algebraic field; the computer obviously does not operate on a closed field because it is restricted to representations of the matrix elements which have limited precision.

The optimum linear MTVF is defined by

$$\underline{\hat{H}}\underline{\hat{R}}_{rr} = \underline{\hat{R}}_{s_d s} \tag{4-5}$$

$$\underline{\hat{H}} = \underline{\hat{R}}_{s_d s}\underline{\hat{R}}_{rr}^{-1}$$

The solution described subsequently employs the spectral decomposition of $\underline{\hat{R}}_{rr}$ given by

$$\underline{\hat{R}}_{rr} = \sum_{i=1}^{kN} \lambda_i \underline{q}_i \underline{q}_i^T \tag{4-6}$$

where the λ_i are the eigenvalues of $\underline{\hat{R}}_{rr}$ and the $\underline{q}_i$ are the associated eigenvectors. The λ_i are ordered so that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{kN} \geq 0. \tag{4-7}$$

Solving for $\underline{\hat{H}}$ using Equation (4-4) requires computing the inverse of $\underline{\hat{R}}_{rr}$. The inverse of $\underline{\hat{R}}_{rr}$ can be obtained from Equation (4-6) and is given by

$$\underline{\hat{R}}_{rr}^{-1} = \sum_{i=1}^{kN} \frac{1}{\lambda_i} \underline{q}_i \underline{q}_i^T \tag{4-8}$$

If some of the λ_i are zero, the inverse given by Equation (4-8) will be undefined. Assuming that there are r nonzero eigenvalues (i.e., the matrix $\underline{\hat{R}}_{rr}$ has rank r), a pseudoinverse may be defined as

$$\underline{\hat{R}}_{rr}^{+} = \sum_{i=1}^{r} \frac{1}{\lambda_i} \underline{q}_i \underline{q}_i^T \tag{4-9}$$

By replacing $\underline{\hat{R}}_{rr}^{-1}$ in Equation (4-5) with the $\underline{\hat{R}}_{rr}^{+}$ of Equation (4-9), a bounded solution for $\underline{\hat{H}}$ can be computed. In practice, however, the problems described in the two previous sections may still cause unacceptable errors in the estimate of $\underline{\hat{H}}$. These errors would cause $\underline{\hat{H}}$ to differ from the theoretically optimum $\underline{\hat{H}}$. If the suboptimum $\underline{\hat{H}}$ were used to filter the received data, large oscillations or undesirable signals could appear in the filter output. The pseudoinverse can be improved by setting small eigenvalues to zero and disregarding the contributions of their associated eigenvectors to the solution. This inverse is called the l-truncated pseudoinverse (Sullivan 1984) and is defined by

$$\underline{\hat{R}}_{rr_l}^{+} \equiv \sum_{i=1}^{l} \frac{1}{\lambda_i} \underline{q}_i \underline{q}_i^T \tag{4-10}$$

The difficulty lies in determining l. Sullivan used the singular value decomposition and the mean-square error to define a criterion for determining

the optimum number of eigenvalues, l_{opt}, to keep for computing the inverse. The l-truncated pseudoinverse was used to extrapolate a signal from m dimensions to n ($n > m$) and assumed additive and uncorrelated zero mean noise. The criterion derived here is for a general class of optimum linear filter equations and accounts for any perturbation to the assumed model for $\underline{R}_{rr}$.

The determination of l_{opt} is now derived by considering the effects of a perturbation in $\hat{\underline{R}}_{rr}$ on the mean-square error. Consider the following expression for the mean-square error:

$$\varsigma^2 = \mathrm{tr}\left[\hat{\underline{R}}_{s_d s_d} + \hat{\underline{H}}\hat{\underline{R}}_{rr}\hat{\underline{H}}^T - 2\hat{\underline{R}}_{s_d s}\hat{\underline{H}}^T\right] \tag{4-11}$$

The filter $\hat{\underline{H}}$ which will be used to evaluate ς^2 is

$$\hat{\underline{H}} = \hat{\underline{R}}_{s_d s}\hat{\underline{R}}^{+}_{rr_t} \tag{4-12}$$

where $\hat{\underline{R}}^{+}_{rr_t}$ is the l-truncated pseudoinverse of $\hat{\underline{R}}_{rr}$ as described in Equation (4-9). Let $\hat{\underline{R}}'_{rr}$ be the estimate using the received signal and noise waveforms and let $\hat{\underline{R}}_{rr}$ be the estimate derived from the assumption $\hat{\underline{R}}_{rr} = \hat{\underline{R}}_{ss} + \hat{\underline{R}}_{nn}$. If $\hat{\underline{R}}_{rr} = \hat{\underline{R}}'_{rr}$, then

$$\varsigma^2 = \mathrm{tr}\left[\hat{\underline{R}}_{s_d s_d} - \hat{\underline{R}}_{s_d s}\hat{\underline{H}}^T\right] \tag{4-13}$$

This is the theoretical mean-square error. If the assumptions, models, and estimates were correct, Equation (4-13) would be the exact mean-square error.

Assume now that $\hat{\underline{R}}_{rr} \neq \hat{\underline{R}}'_{rr}$. In this case the mean-square error is

$$\varsigma^2 = \mathrm{tr}\left[\hat{\underline{R}}_{s_d s_d} + \hat{\underline{H}}\hat{\underline{R}}'_{rr}\hat{\underline{H}}^T - 2\hat{\underline{R}}_{s_d s}\hat{\underline{H}}^T\right] \tag{4-14}$$

where

$$\hat{\underline{R}}'_{rr} = \hat{\underline{R}}_{rr} + \underline{P}_{rr} \tag{4-15}$$

in which $\underline{P}_{rr}$ represents a perturbation on the matrix $\hat{\underline{R}}_{rr}$.

Rewriting Equation (4-14) gives

$$\varsigma^2 = \mathrm{tr}\left[\hat{\underline{R}}_{s_d s_d} + \hat{\underline{H}}\hat{\underline{R}}_{rr}\hat{\underline{H}}^T - 2\hat{\underline{R}}_{s_d s}\hat{\underline{H}}^T + \hat{\underline{H}}\underline{P}_{rr}\hat{\underline{H}}^T\right] \tag{4-16}$$

Using Equation (4-13), this simplifies to

$$\varsigma^2 = \mathrm{tr}\left[\hat{\underline{R}}_{s_d s_d} - \hat{\underline{R}}_{s_d s}\hat{\underline{H}}^T + \hat{\underline{H}}\underline{P}_{rr}\hat{\underline{H}}^T\right] \tag{4-17}$$

Inserting the expression for $\underline{H}$ into Equation (4-17) gives

$$\varsigma^2 = \mathrm{tr}\left[\hat{\underline{R}}_{s_d s_d}\right] - \mathrm{tr}\left[\hat{\underline{R}}_{s_d s}\hat{\underline{R}}^{+}_{rr_l}\hat{\underline{R}}^{T}_{s_d s}\right] + \mathrm{tr}\left[\hat{\underline{R}}_{s_d s}\hat{\underline{R}}^{+}_{rr_l}\underline{P}_{rr}\hat{\underline{R}}^{+}_{rr_l}\hat{\underline{R}}^{T}_{s_d s}\right] \tag{4-18}$$

Note that although $\underline{P}_{rr}$ is a symmetric matrix, it may have negative eigenvalues.

The minimum norm least square (MNLS) criterion is as follows:

(1) *insert the* $\hat{\underline{R}}^{+}_{rr_l}$ *of Equation* (4-9) *into the expression for the mean-square* error given by Equation (4-18),

(2) evaluate Equation (4-18) for all possible values of l,

(3) determine l_{opt} as that value of l for which the mean-square error is a minimum,

(4) set the remaining $kN - l_{opt}$ eigenvalues of $\hat{\underline{R}}_{rr}$ equal to zero, and

(5) compute the MNLS estimate of $\underline{H}$ using $\hat{\underline{R}}^{+}_{rr_{l_{opt}}}$.

The MNLS algorithm just described could be implemented by computing a new $\hat{\underline{H}}$ for each value of l and evaluating Equation (4-18), but this method would be very inefficient and, as a result, time consuming. There is a good deal of redundant information available in ς^2 computed using $\hat{\underline{R}}^{+}_{rr_{l-1}}$ which can

be used to compute ς^2 for $\hat{\underline{R}}^{+}_{rr_l}$. This suggests a recursive algorithm for calculating $\varsigma^2(l)$, which is the mean-square error as a function of l. The recursive algorithm is derived in Appendix B. The computational complexity of the eigenanalysis along with the computational complexity of two MNLS algorithms are compared in Table 4-1. The simplistic MNLS refers to computing the mean-square error using Equation (4-18) directly for each new eigenvalue. The table clearly shows that the computational burden of the simplistic MNLS algorithm dominates that of the eigenanalysis. The recursive MNLS algorithm requires fewer computations than the eigenanalysis. The computer time is dominated by the MNLS computation using the simplistic algorithm, but by the eigenanalysis (which must be done anyway) using the recursive algorithm. Using the recursive algorithm can therefore save a great deal of time and requires only moderately additional overhead to the eigenanalysis.

There is one problem which arises in this implementation which must be addressed. The mean-square error can be written as a sum of two parts: (1) the optimum matrix and (2) the perturbation matrix. Consider again equation (6.4.16) rewritten to emphasize these parts.

$$\varsigma^2 = \mathrm{tr}\left[\hat{\underline{R}}_{s_d s_d} - \hat{\underline{R}}_{s_d s}\hat{\underline{H}}^{T}\right] + \mathrm{tr}\left[\hat{\underline{H}}\underline{P}_{rr}\hat{\underline{H}}^{T}\right] \qquad (4\text{-}19)$$

The part due to the optimum matrix will decrease with increasing l and approach zero. Theoretically, it could never be less than zero, but may be slightly negative due to the variance of the estimates or the ill-conditioned nature of $\hat{\underline{R}}_{rr}$ (used in obtaining $\hat{\underline{H}}$). Since this cannot be allowed, special attention is given to this case when determining l_{opt} in software. The part due to the perturbation matrix may also be negative because $\underline{P}_{rr}$ is not necessarily positive definite. If such is the case, a slightly optimistic mean-square error is obtained because the second part subtracts from the first. If the second part is negative enough, the total mean-square error may become negative and this is clearly impossible. This situation, although theoretically impossible, can be explained. The optimum linear filter was derived under certain assumptions. If these assumptions are invalid, the derived equations for the mean-square error will also be invalid and as a result, the second part of ς^2 due to $\underline{P}_{rr}$ can be negative. If the second part is negative for all values of l, then it is highly probable that the assumptions, models, and estimates used to obtain $\hat{\underline{H}}$ are incorrect. This situation is also detected in software.

Table 4-1 Computational Complexities of Eigenanalysis and MNLS (Minimum Norm Least Squares) Algorithms

Algorithm	Computations
Simplistic MNLS	$O\left[(kN)^4\right]$
Eigenanalysis	$O\left[(kN)^3\right]$
Recursive MNLS	$O(k^2N^3)$

It is much more likely that the second part will be negative when only a small number of eigenvalues are kept ($l < l_{opt} << kN$). This is because a suboptimal filter is being used to evaluate the mean-square error. If the suboptimal filter were used, the filter might not minimize the mean-square error as designed, but actually increase it by allowing signals larger in the output than the input. As more eigenvalues are kept, the filter approaches the optimum filter and the second part becomes positive. In order to avoid obtaining an optimistic mean-square error when the second part is negative, the absolute value of the second term is used instead. The minimum mean-square error will be found where the magnitude of the second term is near zero. As the number of eigenvalues kept approaches kN, the mean-square error may become very large due to the problems discussed in the first two sections. Practically, this is because relatively small eigenvalues, when inverted, cause the noise term due to $\underline{P}_{rr}$ to increase much faster than the theoretical mean-square error decreases. There will be a value of l, designated l_{opt}, for which the mean-square error will be a minimum. The largest l_{opt} eigenvalues are kept and the remaining $kN - l_{opt}$ eigenvalues are set to zero. The resulting $\hat{\underline{R}}^{+}_{rr_t}$ is then used to obtain the MNLS solution for $\hat{\underline{H}}$.

4.4. Simulation Test

In this section, the MNLS criterion is tested on data which simulate the perturbations described in the previous section. Gaussian signal and noise data records with zero means and prespecified covariance matrices were generated according to Fukunaga (1972). The covariance matrices for the signal and noise data records were chosen according to Standard Data Sets 1 and 2 in Fukunaga (1972 pp. 46-47). Two hundred signal records and four hundred noise records were generated. Each data record contained eight sample points. A set of gaussian received data records were simulated by adding one hundred noise records to one hundred signal records. A second channel was simulated by adding a different set of one hundred noise records to the same one hundred signal records used to simulate the first channel. The SNR was set to -6 dB by scaling the noise records appropriately. These simulated received data records are referred to as the testing set. Two different MTVFs were designed using different training sets (the data records used to estimate the signal and noise statistics). The first MTVF used identical training and testing sets. The second MTVF used a training set consisting of the remaining signal and noise records not used to make up the simulated data but generated from the same covariance matrices. In practice, the statistics of the signal and noise processes

are seldom known exactly but must be estimated. The VEP problem, as an example, requires the indirect estimation of the multichannel cross-correlation matrices for both the poststimulus EEG and VEP. This is the difference simulated by the two MTVF designs.

The MTVF designed for each of the two training sets was first computed by keeping only the largest eigenvalue (and its associated eigenvector) in the pseudoinverse. The theoretical, actual, and minimum norm mean-square errors were then calculated. This was repeated sixteen times, adding the next largest eigenvalue to the pseudoinverse at each step and recalculating the mean-square errors. The actual mean-square error was computed using the actual and estimated signal records. The MNLS criterion was then checked by comparing the theoretical and actual mean-square errors to determine whether the theoretical or the minimum norm mean-square error was a better indicator of the actual error.

The MTVF designed using identical training and testing sets performed well considering the low input SNR and similarity between the signal and noise processes. Table 4-2 summarizes the theoretical, actual, and minimum norm mean-square errors. The MNLS criterion suggested retaining all sixteen eigenvalues as did the theoretical mean-square error. The actual mean-square error was lowest when only fifteen eigenvalues were kept to compute the pseudoinverse. In this case, both criterion were close to predicting the optimum number of eigenvalues to keep.

The MTVF designed using different training and testing sets is more interesting. These results are summarized in Table 4-3. The MNLS criterion suggests retaining only ten eigenvalues. The theoretical mean-square error is lowest when keeping all sixteen eigenvalues. The lowest actual mean-square error was obtained by keeping twelve eigenvalues. Retaining more eigenvalues causes the actual mean-square error to increase even though the theoretical error continues to decrease. The theoretical mean-square error will always continue to decrease as more eigenvalues are retained for computing th pseudoinverse. The MNLS criterion resulted in an actual mean-square error which was the second lowest. Figure 4-2 depicts these results more clearly. Note that initially, the mean-square errors are very high since only one eigenvalue is kept. The mean-square errors converge when keeping four to six eigenvalues and begin to diverge when keeping more than seven. After seven eigenvalues, the theoretical mean-square error continues to decrease but much more slowly. The theoretical mean-square error converges to the optimum mean-square error for these data. The minimum norm criterion results in a

Table 4-2 Minimum Norm Simulation Test Using Identical Training and Testing Sets

Eigenvalues Kept	Theoretical MSE	Actual MSE	MN MSE
16	0.384	0.387	0.404
15	0.389	0.384	0.417
14	0.408	0.413	0.428
13	0.413	0.415	0.433
12	0.415	0.415	0.436
11	0.416	0.417	0.437
10	0.424	0.424	0.448
9	0.457	0.442	0.489
8	0.598	0.562	0.614
7	0.623	0.584	0.631
6	0.636	0.604	0.642
5	0.686	0.654	0.701
4	0.786	0.752	0.801
3	0.801	0.768	0.818
2	0.811	0.787	0.826
1	0.950	0.962	0.951

Table 4-3 Minimum Norm Simulation Test Using Different Training and Testing Sets

Eigenvalues Kept	Theoretical MSE	Actual MSE	MN MSE
16	0.359	0.490	0.547
15	0.363	0.492	0.534
14	0.385	0.489	0.514
13	0.391	0.487	0.513
12	0.394	0.481	0.503
11	0.401	0.486	0.498
10	0.405	0.484	0.495
9	0.410	0.487	0.499
8	0.411	0.507	0.521
7	0.500	0.563	0.587
6	0.550	0.609	0.644
5	0.722	0.742	0.724
4	0.837	0.845	0.858
3	0.864	0.835	0.895
2	0.894	0.858	0.914
1	0.899	0.877	0.918

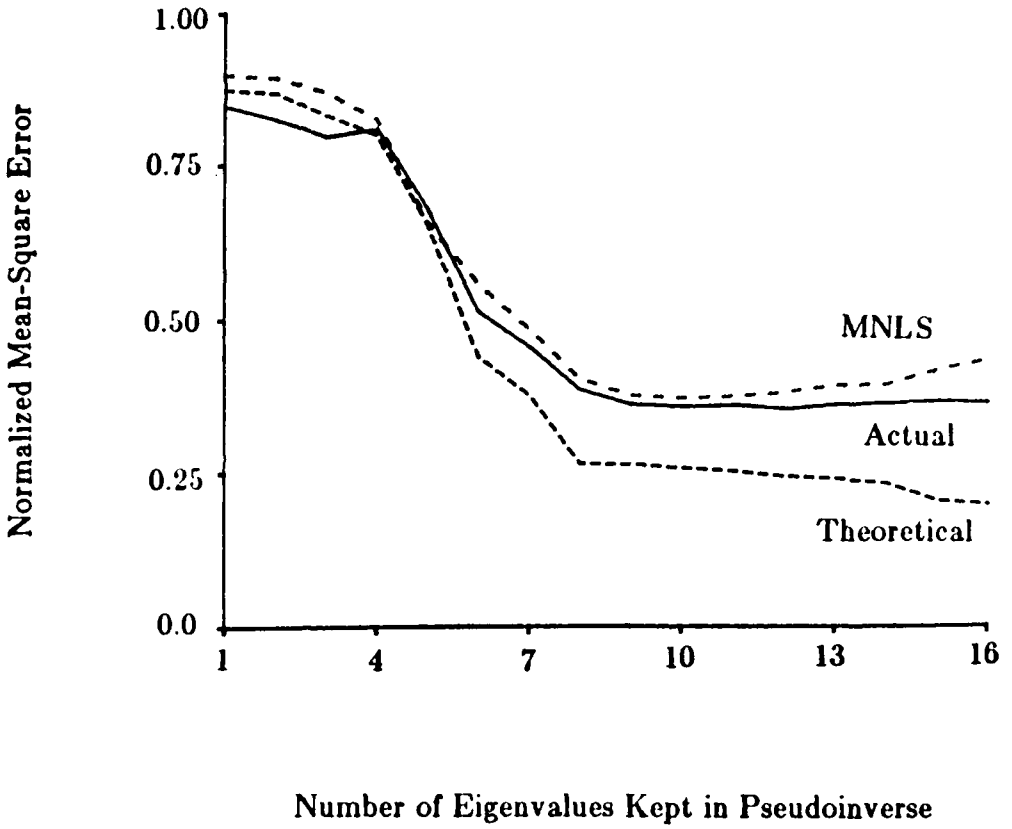


Figure 4-2 Mean-Square Error Comparisons Using Different Training and Testing Sets

mean-square error which begins to increase again after ten eigenvalues. The actual mean-square error appears to be bounded pessimistically by the minimum norm error and optimistically by the theoretical error. This result confirms that using more eigenvalues than necessary can increase the actual mean-square error even though the theoretical error continues to decrease. The MNLS criterion suggests a near optimum number of eigenvalues for designing the MTVF when the training and testing data sets are different. This is typically the case for practical optimum filtering problems and is certainly the case for VEP estimation. The implications of this result will be seen later when the dimensionality increases and the differences between the training and testing sets are magnified. The output of the MTVF will be shown to oscillate drastically when certain suboptimum numbers of eigenvalues are retained.

4.5. Evoked Potential Data Tests

Scalp-recorded brain responses following checkerboard patterned stimulation of the lower visual field of human subjects were used to test the performance of the optimum MTVF. Prestimulus on-going EEG data were also recorded for the purposes of artifact rejection and estimation of the noise cross-correlation matrix. The EP was modeled as a sum of raised cosine components with randomly varying amplitudes and latencies. Random signal component data records were generated and used in an initial simulation to establish an optimistic bound on the two channel performance. The parameters used to generate the simulated EPs were obtained from a latency corrected average (LCA) of the scalp-recorded responses at electrode Pz (McGillem 1977 and Aunon 1979). Table 4-4 summarizes the LCA results at electrode Pz and Table 4-5 at electrode Cz. These signals were added to the on-going EEG data records recorded at electrode Pz prior to stimulus presentation. The noise records were scaled to provide a desired input SNR. A second channel of simulated scalp-recordings was obtained by adding the simulated VEP signals (scaled by a factor of 0.8) to the prestimulus EEG recorded at electrode Cz. Using a uniform scale factor for each component is not a requirement of the random signal model, but is convenient and should provide an optimistic bound on the filter performance when the filter is applied to the human VEP data. Example waveforms (-6 dB SNR) are plotted in Figure 4-3. They appear to resemble scalp-recorded responses. Depending on the latencies of larger EEG components, the signal peaks are either obscured or exaggerated.

Table 4-4 Lower Checkerboard Latency Corrected Average Results at Electrode Pz (Subject #5)

Peak	Latency (ms)	Latency St. Dev. (ms)	Amplitude (uV)
1	74.14	7.41	6.28
2	111.60	6.65	-13.07
3	146.87	8.99	3.11
4	170.00	7.81	-2.22
5	202.95	11.32	8.42

Table 4-5 Lower Checkerboard Latency Corrected Average Results at Electrode Cz (Subject #5)

Peak	Latency (ms)	Latency St. Dev. (ms)	Amplitude (uV)
1	73.31	8.76	5.17
2	113.79	10.97	-12.42
3	166.40	5.72	-3.19
4	199.58	11.72	11.16

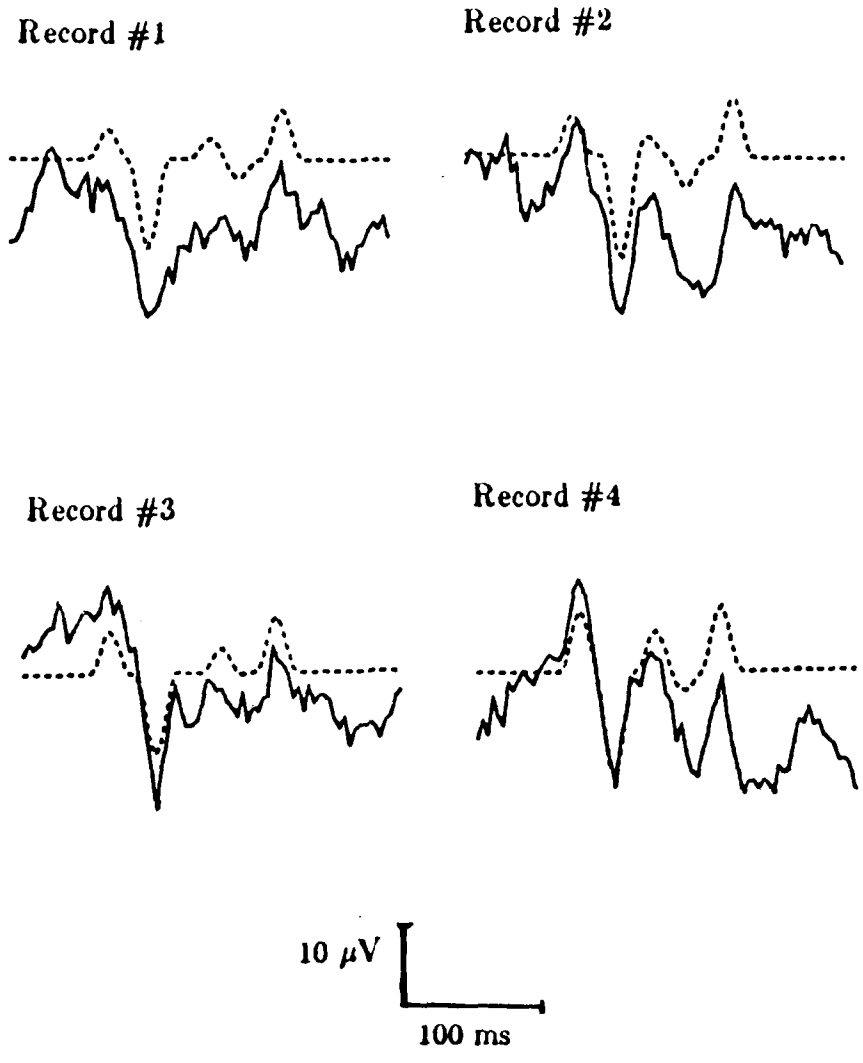


Figure 4-3 Example Waveforms (solid) for Simulated Evoked Potential Signal (dash) in Prestimulus Electroencephalogram (Electrode Pz) (-6 dB SNR)

Single and two channel filters were designed using the simulated scalp-recordings. The cross-correlation matrix of the VEP signals was derived from the random signal component model under the assumptions of independent components and gaussian distributions for both the amplitudes and latencies. The single and two channel filters were designed according to the method outlined in the implementation subsection. The performance criteria are summarized in Table 4-6. Figure 4-4 depicts these performance improvements vs. the input SNR. Considerable improvements are realized ranging from a 26% (-9 dB) to a 30% (0 dB) reduction in the actual MSE using the two channel filter. The output SNR is improved by approximately 2 dB. Note that even at low input SNRs, the two channel filter provides considerable performance improvements.

Example waveforms (-6 dB input SNR) reflecting the mean-square error (MSE), noise reduction factor (NRF), and bias factor (BF) of the optimum filters are plotted in Figures 4-5 through 4-7. Examples of the single and two channel filter outputs for simulated VEP plus human EEG input are plotted in Figure 4-5. The two channel filter reduces the MSE in the signal estimate especially at latencies where the SNR is small, such as near the second positive and negative signal peaks. Examples of the noise reduction improvements realizable using the two channel filter are plotted in Figure 4-6. Note especially Record #1 in which a component in the output of the single channel filter at 200 ms is suppressed by the two channel filter. The filter bias is depicted in Figure 4-7. Noiseless random signals were passed through the optimum filters to determine the degree of signal distortion. As predicted by the performance criteria, there is little bias in either filter output.

The performances of the single and two channel filters on the human scalp-recorded data are now considered. Figure 4-8 depicts the average VEPs to lower checkerboard stimulation for Subject #5 recorded at electrodes Pz and Cz. Note that the on-going EEG in Figure 4-8a is very nearly zero mean and that the average VEPs are quite similar in both electrodes. The average VEP is a transient signal with a distinct onset. Because the VEP also varies randomly from one response to the next, the scalp-recorded random process must be considered nonstationary. Example scalp-recorded responses are plotted in Figure 4-9. The VEP is buried in the on-going EEG and only a trained eye could detect the locations of the larger peaks in the VEP and only then by knowing the time of stimulus presentation. The input SNR is approximately -7 dB (as determined from the random signal model and the prestimulus EEG).

Table 4-6 Performance Comparisons vs. Input Signal-to-Noise Ratio for Simulated Evoked Potential Signal in Prestimulus Electroencephalogram. Abbreviations Used are MSE (Mean-Square Error), MNLS (Minimum Norm Least Square), BF (Bias Factor), NRF (Noise Reduction Factor), SNR_I (Input Signal-to-Noise Ratio), and SNR_O (Output Signal-to-Noise Ratio). MSE and BF are Normalized.

Criterion	Number of Channels	Input SNR (dB)			
		0	-3	-6	-9
Theoretical MSE	Single	0.037	0.056	0.085	0.125
	Two	0.025	0.039	0.060	0.091
Actual MSE	Single	0.037	0.055	0.084	0.125
	Two	0.026	0.040	0.061	0.092
MNLS Error	Single	0.050	0.061	0.086	0.138
	Two	0.051	0.057	0.077	0.104
BF	Single	0.016	0.024	0.037	0.054
	Two	0.011	0.016	0.026	0.042
NRF (dB)	Single	-16.635	-17.909	-19.130	-20.482
	Two	-18.321	-19.518	-20.745	-22.066
SNR_O (dB)	Single	16.372	14.508	12.506	10.536
	Two	18.143	16.246	14.320	12.407

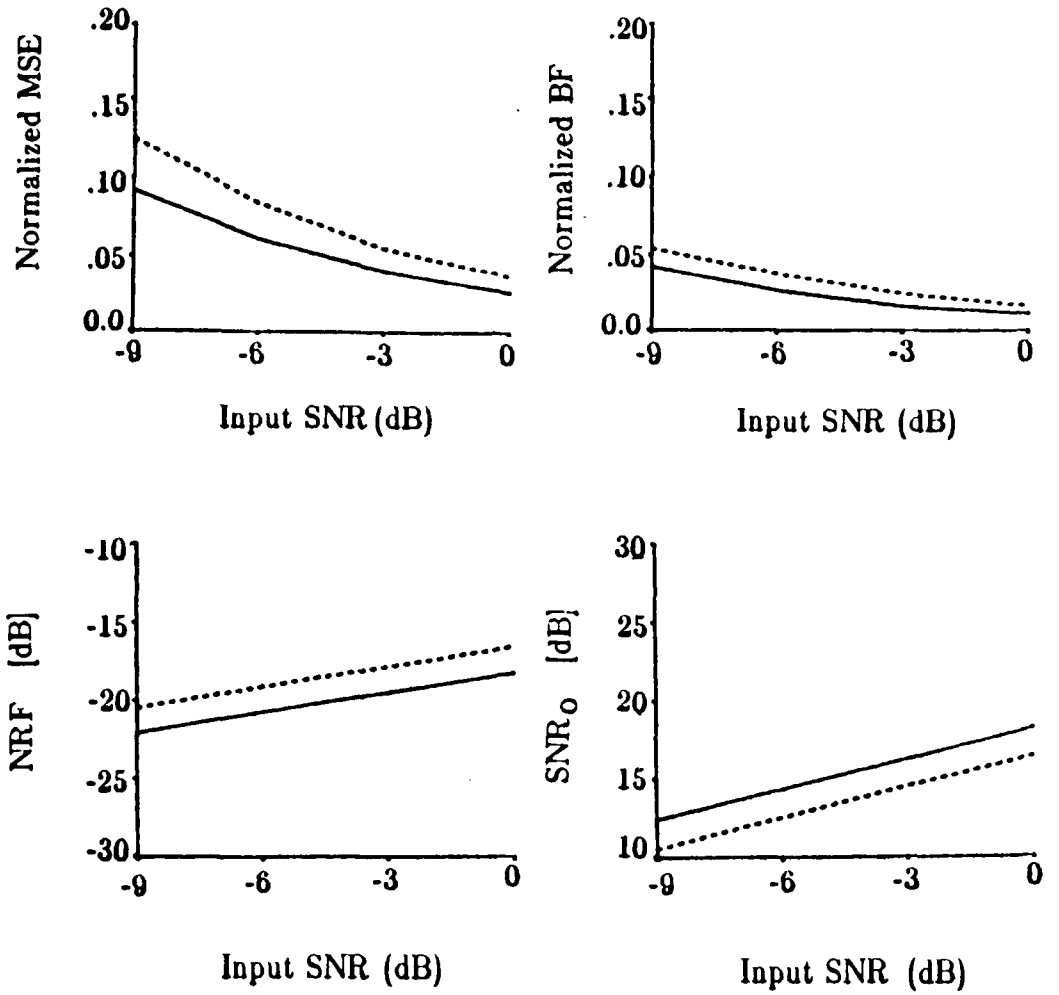


Figure 4-4 Performance Comparison vs. Signal-to-Noise Ratio for Single (dash) and Two Channel (solid) Filters on Simulated Evoked Potential Data. Abbreviations Used are Actual MSE (Mean-Square Error), BF (Bias Factor), NRF (Noise Reduction Factor), and SNR_O (Output Signal-to-Noise Ratio).

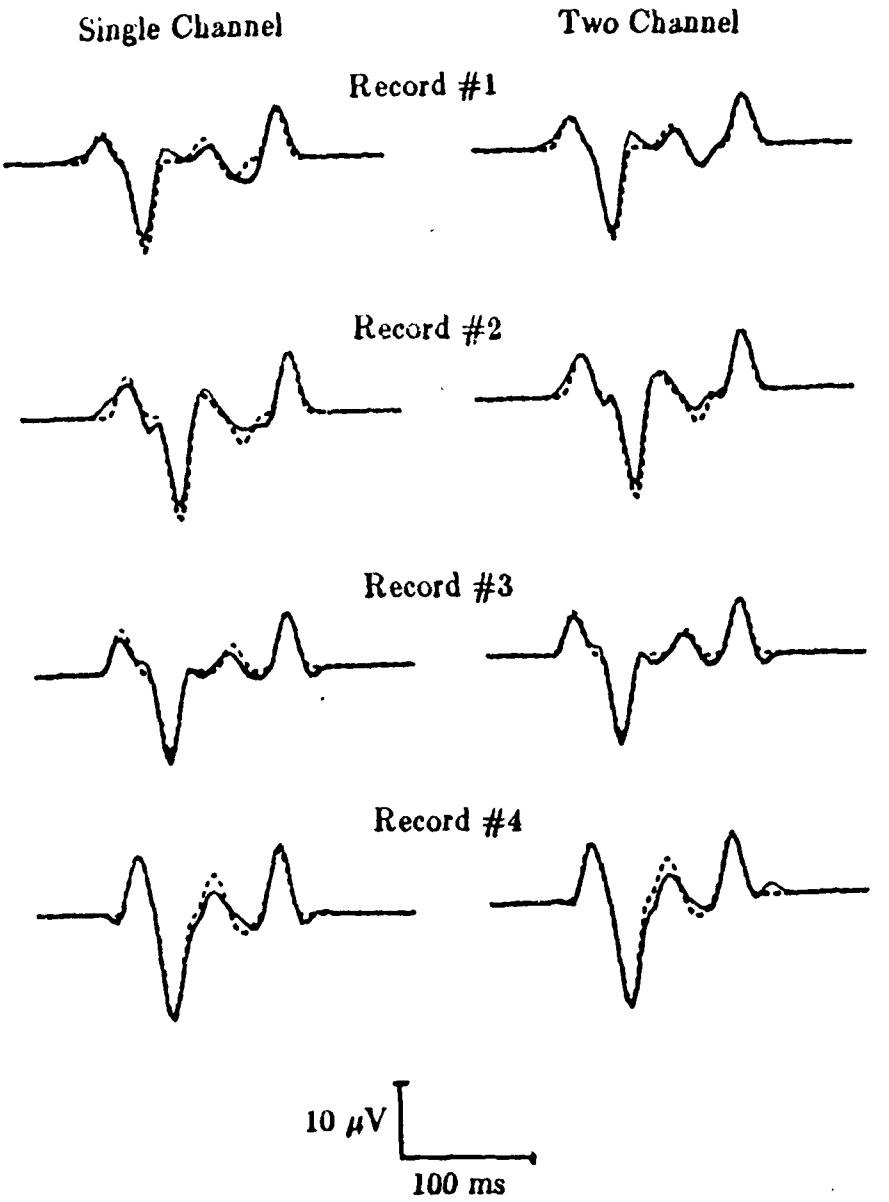


Figure 4-5 Examples of Single and Two Channel Filter Output (solid) for Simulated Evoked Potential Signal (dash) Plus Electroencephalogram Input (-6 dB SNR)

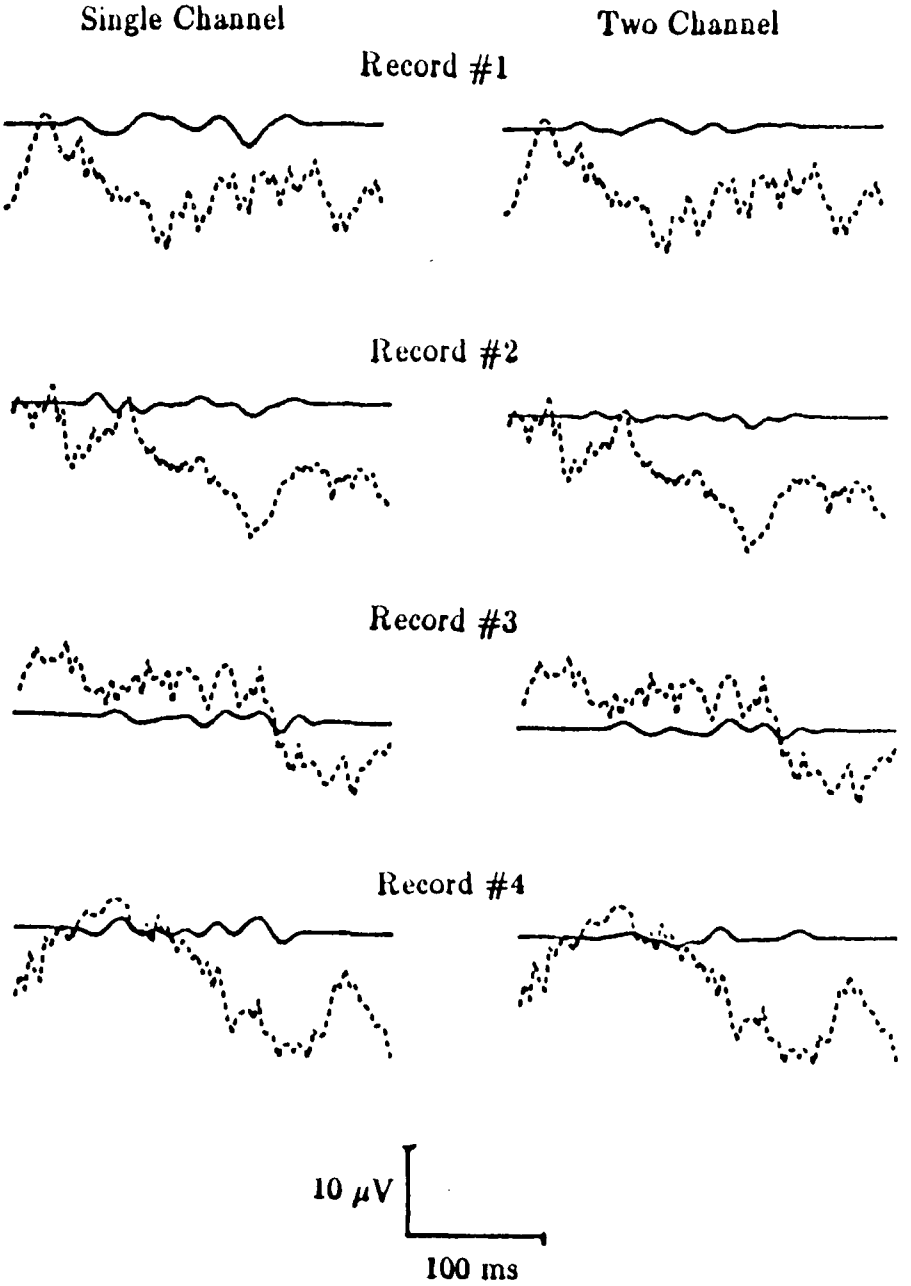


Figure 4-6 Examples of Single and Two Channel Filter Output (solid) for Prestimulus Electroencephalogram Only Input (dash)

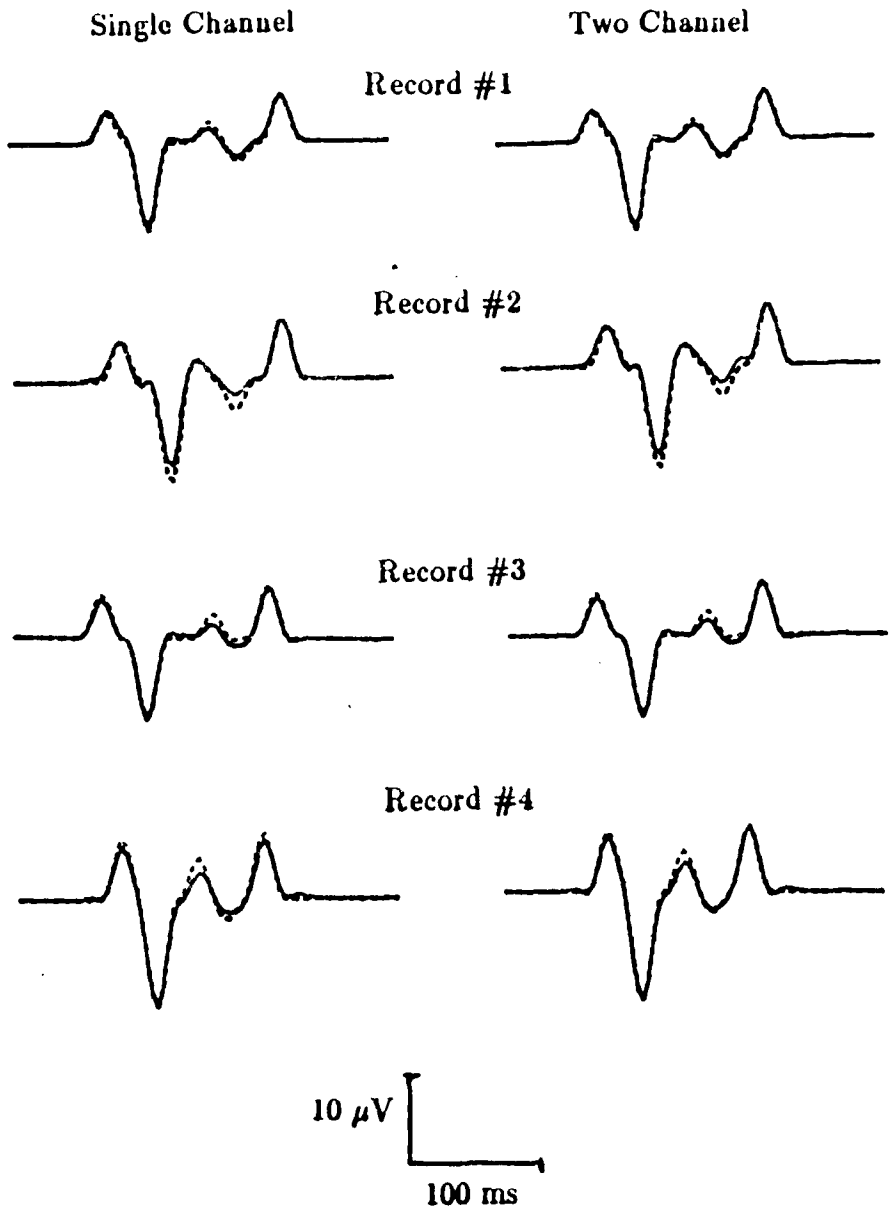


Figure 4-7 Examples of Single and Two Channel Filter Output (solid) for Simulated Evoked Potential Signal Only Input (dash)

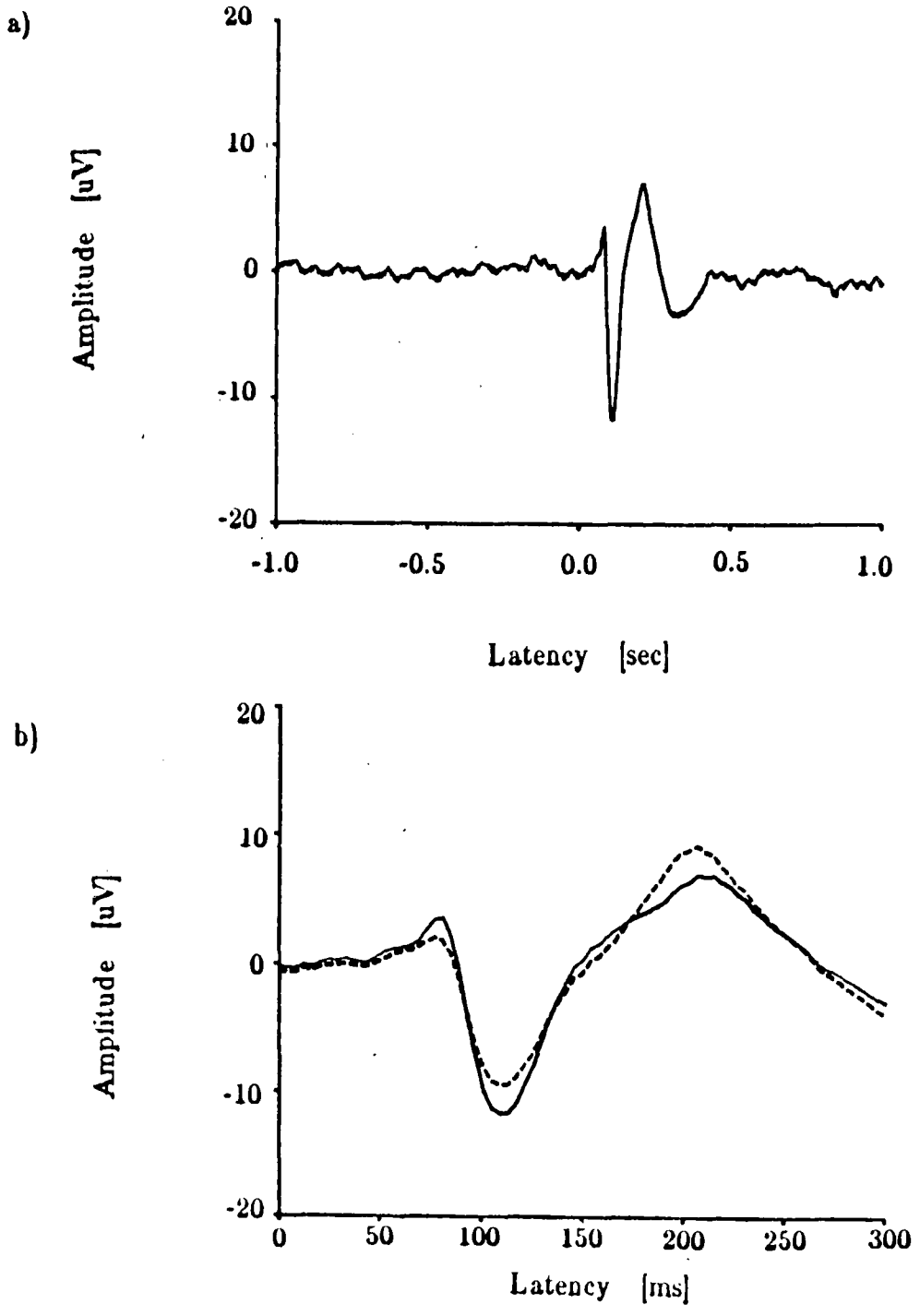


Figure 4-8 Average Visual Evoked Potentials to Lower Checkerboard Stimulus (Subject #5). a) Electrode Pz Including Prestimulus and b) Electrodes Pz (solid) and Cz (dash) Following Stimulus.

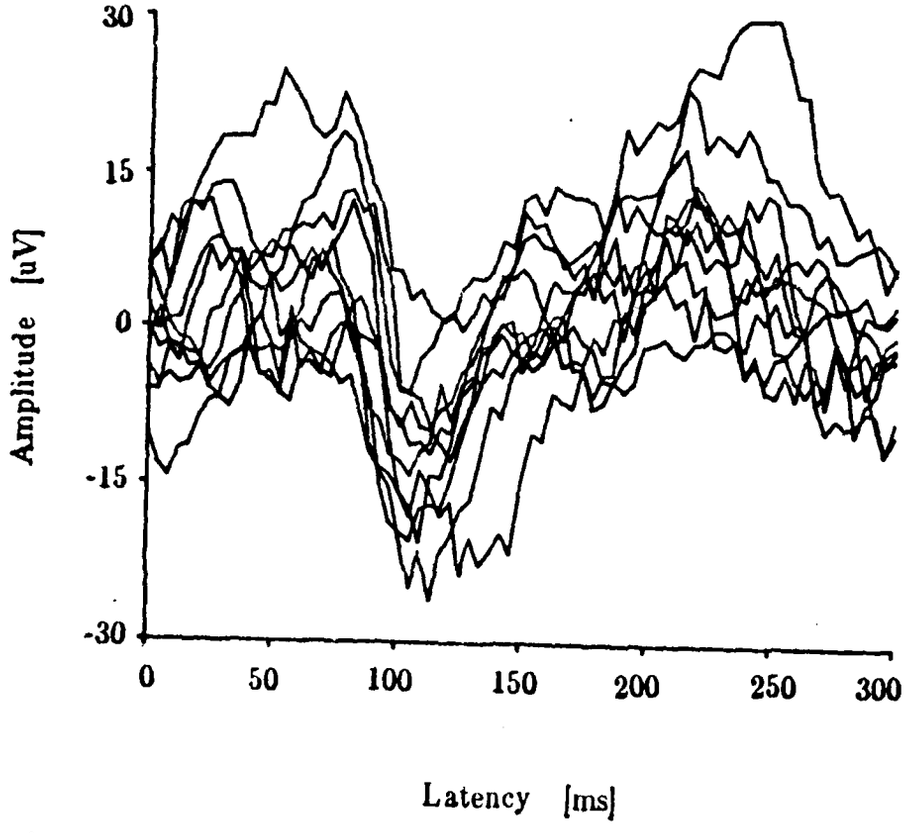


Figure 4-9 First Ten Superimposed Scalp-Recorded Responses at Electrode Pz Following Lower Checkerboard Stimulus (Subject #5)

Single and two channel optimum filters were designed according to the method outlined in the implementation subsection. The filters were designed to estimate the VEP in electrode Pz. The VEP cross-correlation matrix was obtained using the random signal model. The necessary statistics of the individual components of the VEP for subject #5 were determined from the LCA. These statistics were summarized in Table 4-4 for electrode Pz and Table 4-5 for electrode Cz. The lower checkerboard average VEP is typically largest in electrodes Oz and Pz (Jeffreys 1971). Electrodes Pz and Cz were chosen to design the two channel filter (estimating the signal in electrode Pz) because the LCA for both of these channels detected a negative component at 160 ms which did not appear in the LCA for electrode Oz. Since it was desired to estimate this component more effectively, the scalp-recorded responses from electrode Cz were chosen instead of those from electrode Oz.

The MNLS criterion was an important factor in obtaining a stable two channel filter for the human VEP data. The cross-correlation matrices used to design the filter may both be slightly different than those for the scalp-recorded responses. The prestimulus EEG may not be a good model of the post-stimulus EEG if the statistics of the EEG change due to the presence of the VEP. The random signal model may also be imperfect. If the VEP and EEG are in fact correlated, the assumptions used to design the filter would be violated and it is possible that the output of the filter would become unstable if the MNLS criterion were not used.

Figure 4-10 summarizes the MNLS error vs. the number of eigenvalues used to compute the pseudoinverse. In the single channel case, the MNLS error decreased rapidly until the 10th eigenvalue (approximately the number of significant eigenvectors required to span the output signal space). After the 10th eigenvalue, the MNLS error remained fairly constant, but slowly began to decrease again after the 50th eigenvalue. The optimum number of eigenvalues to retain was 74 and the MNLS error remained bounded at all times. The two channel results were quite different. Once again, the MNLS error decreased rapidly until the 15th eigenvalue from which point it continued to decrease but at a much slower rate. The optimum number of eigenvalues to keep was 112. After 112 eigenvalues, however, the MNLS error began to increase rapidly. After 130 eigenvalues, the MNLS error was greater than 1.0 suggesting that the output of the two channel filter was highly unstable.

Example waveforms depicting the output of the two channel filter are plotted in Figure 4-11. It is apparent that the individual components in the VEP estimates are varying randomly in both amplitude and latency. This

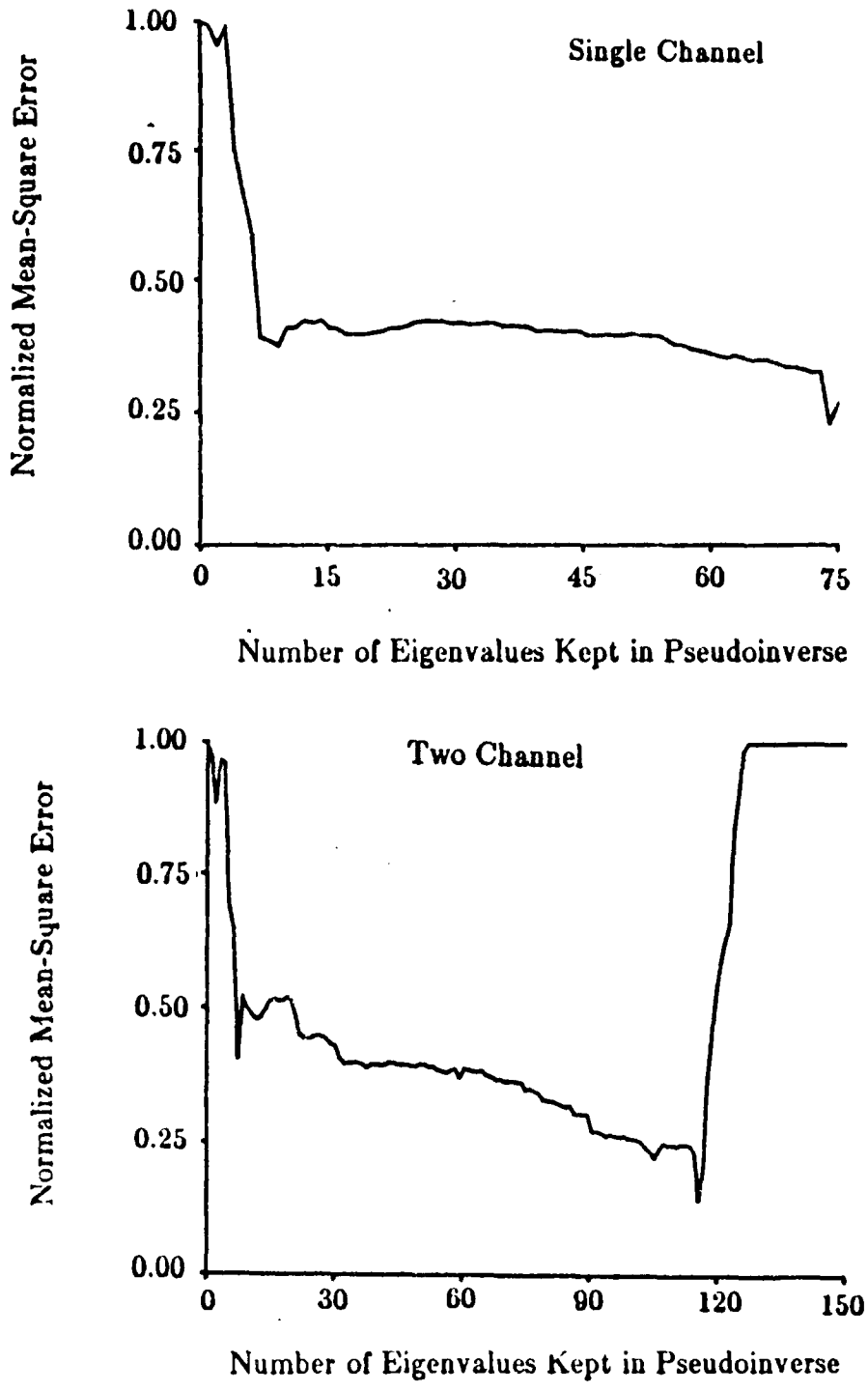


Figure 4-10 Mean-Square Error vs. Number of Eigenvalues Kept in Pseudoinverse for Minimum Norm Criterion Using Lower Checkerboard Responses from Electrodes Pz and Cz (Subject #5)

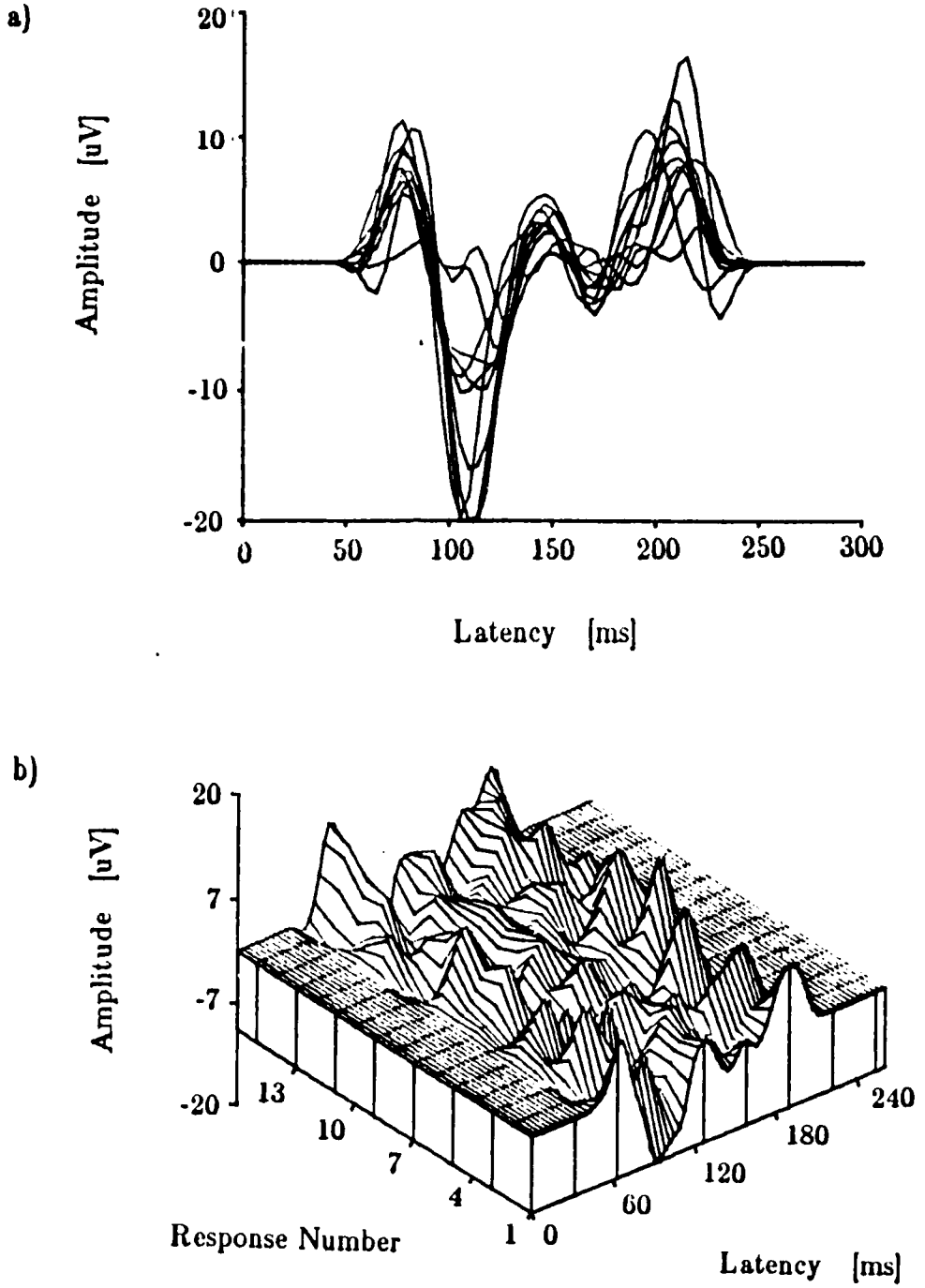


Figure 4-11 Examples of Two Channel Filtered Scalp-Recorded Responses From Electrodes Pz and Cz Following Lower Checkerboard Stimulus (Subject #5). a) Ten Superimposed Filtered VEPs and b) Three Dimensional View Showing Variability with Each Response. Estimating Visual Evoked Potential in Electrode Pz.

suggests that the random signal model is providing the intended result.

The performance criteria for the single and two channel filters are summarized in Table 4-7. The theoretical MSE and MNLS error suggest a reduction in the actual MSE of between 32% and 35% using the two channel filter. The output SNR should improve by more than 2 dB. Examples of the filter outputs and the scalp-recorded inputs are plotted in Figures 4-12 and 4-13. Note in Record #5 that the two channel filter appears to improve the estimate of the third positive peak because of the additional information provided by the larger and narrower peak in electrode Cz. The estimate of the second negative peak is improved in the same manner in Record #9 and Record #26. In Record #9, the single channel estimate looks quite adequate. By examining the response in electrode Cz, however, it appears that the single channel filter has incorrectly located the second negative peak. In electrode Pz, there are two negative deflections in the response near where the second negative VEP component should be. Without additional information, the single channel filter chooses the later deflection. The response in electrode Cz, however, shows only the first deflection and as a result, the two channel filter attempts to place the component nearer to the first deflection. The additional information in the Cz response also helps to place the second negative VEP component in Record #26.

Figures 4-14 and 4-15 depict examples of the noise reduction capabilities of the optimum filter. Note in Records #8 and #10 especially the ability of the two channel filter to suppress EEG components which appear in the output of the single channel filter. It is important that the on-going EEG be reduced as much as possible so that noise components are not construed as VEP.

Examples of the bias associated with the single and two channel filters are plotted in Figure 4-16. Note especially that the two channel filter does not distort the smaller signal components (the second positive and negative components) as much as the single channel filter.

It is important to realize that the estimate is only as good as the signal model. Improved signal models using better physiologically descriptive components may improve the VEP estimates even further. The simple raised cosine model is useful if the peak amplitudes and latencies are more important than the morphology of the VEP. The use of additional channels may or may not improve the VEP estimates. Additional channels may improve the estimates of the eigenvectors (both for the signal output space and the noise space) and thus improve the VEP estimates, but the computational burden and limitations of the computer precision may override the modest gains

Table 4-7 Performance Comparisons Single vs. Two Channel Filter on Lower Checkerboard Visual Evoked Potentials (Subject #5). Abbreviations Used are MSE (Mean-Square Error), MNLSE (Minimum Norm Least Square Error), BF (Bias Factor), NRF (Noise Reduction Factor), and SNR_O (Output Signal-to-Noise Ratio). MSE and BF are Normalized.

Criterion	Single	Two
Theoretical MSE	0.156	0.106
MNLSE	0.227	0.147
BF	0.068	0.048
SNR_I (dB)	-7.756	-7.756
NRF (dB)	-18.312	-20.109
SNR_O (dB)	9.342	11.576

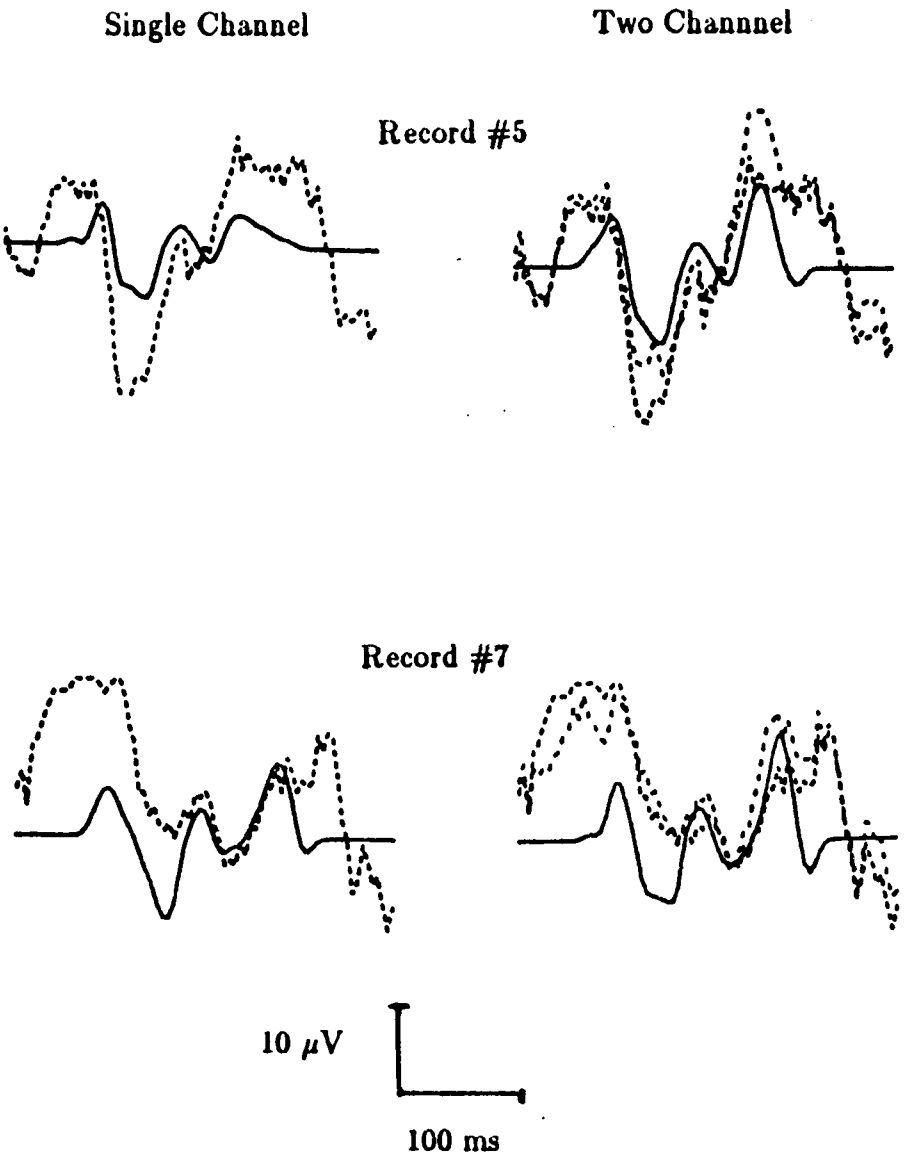


Figure 4-12 Comparison of Single and Two Channel Filter Outputs Using Human Scalp-Recorded Responses Following a Lower Checkerboard Stimulus (Subject #5). Pz (dash), Cz (dash with wider gap), and Filter Output (solid).

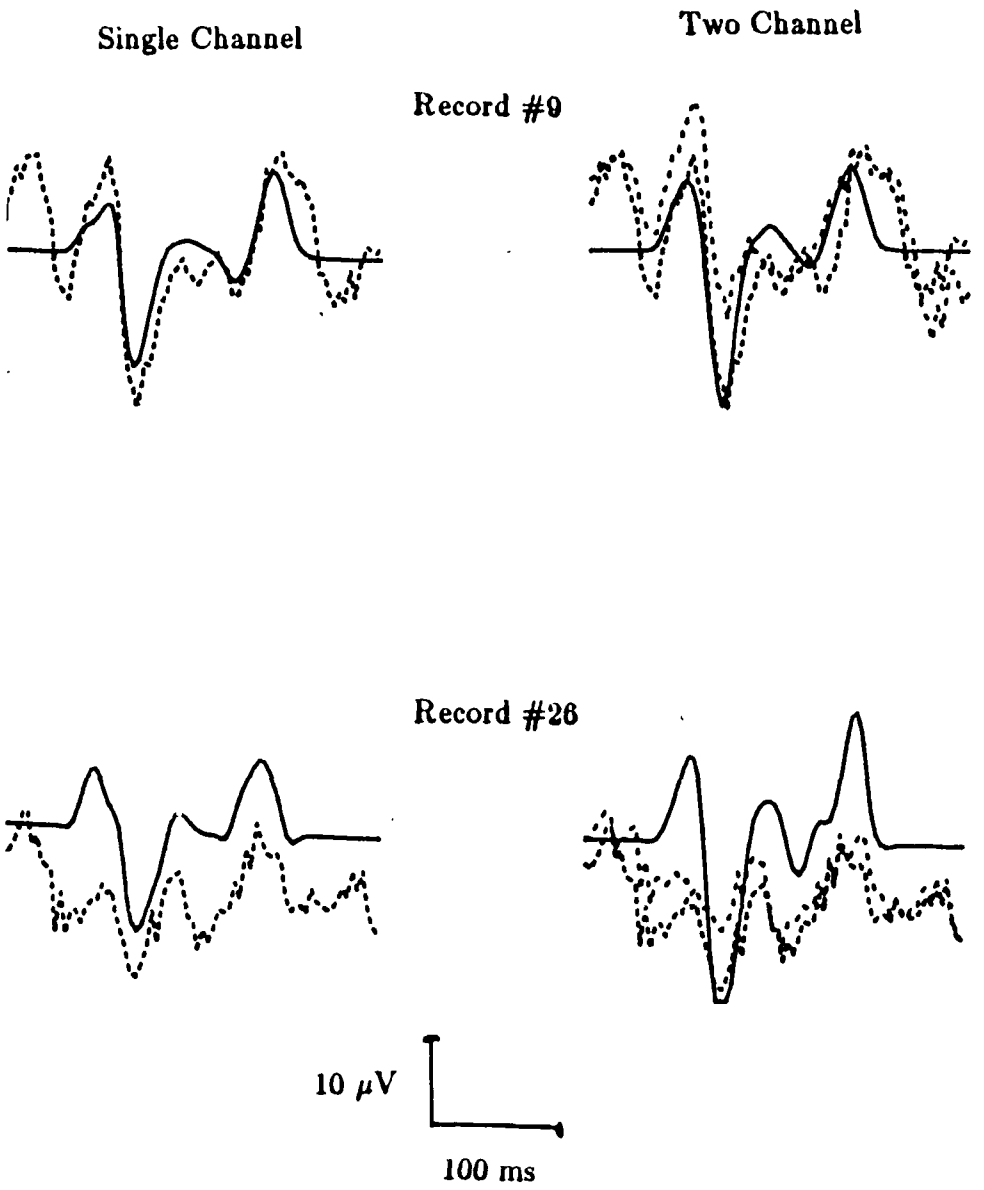


Figure 4-13 Comparison of Single and Two Channel Filter Outputs Using Human Scalp-Recorded Responses Following a Lower Checkerboard Stimulus (Subject #5). Pz (dash), Cz (dash with wider gap), and Filter Output (solid).

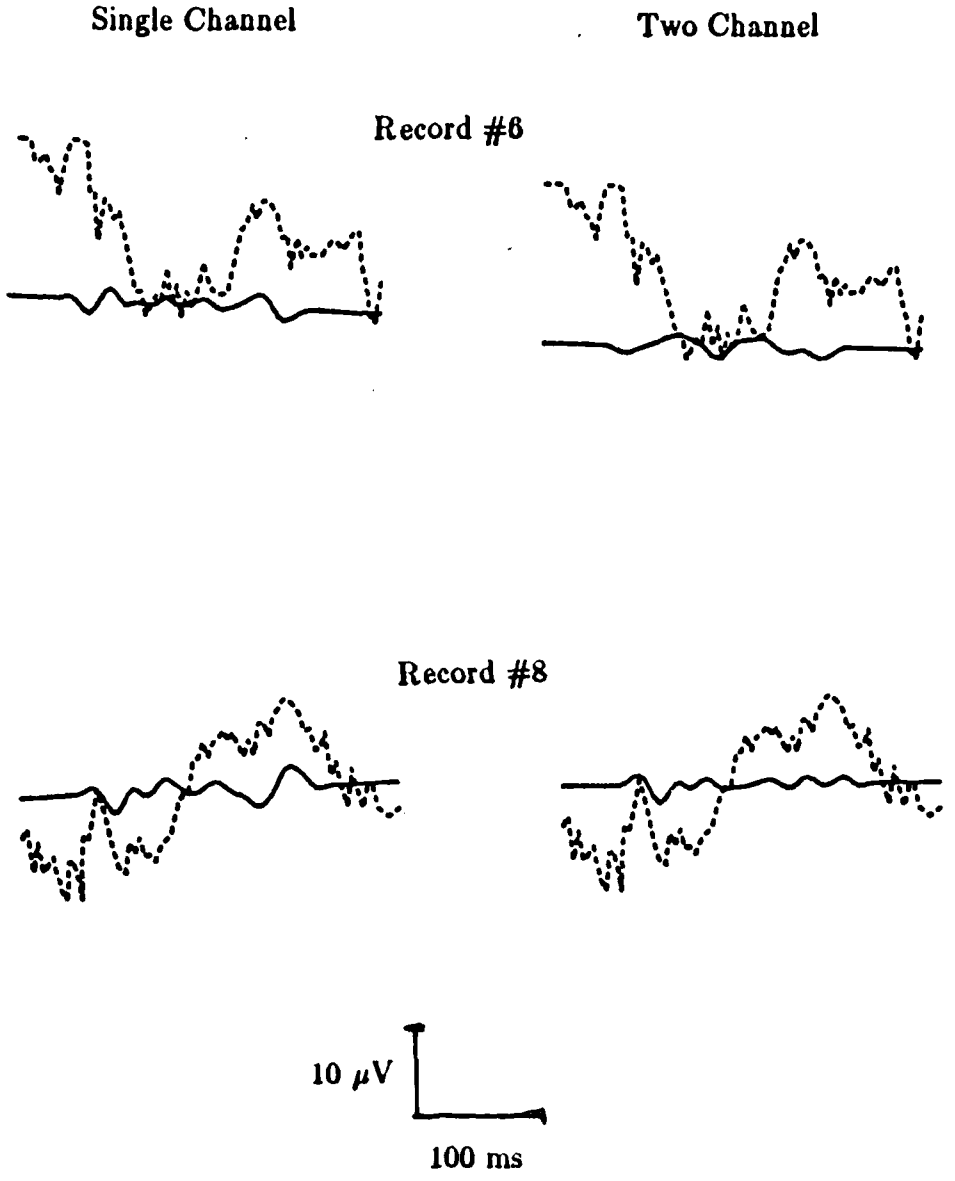


Figure 4-14 Comparison of Single and Two Channel Filter Output Using Prestimulus Electroencephalogram as Input (Subject #5). Pz (dash) and Filter Output (solid).

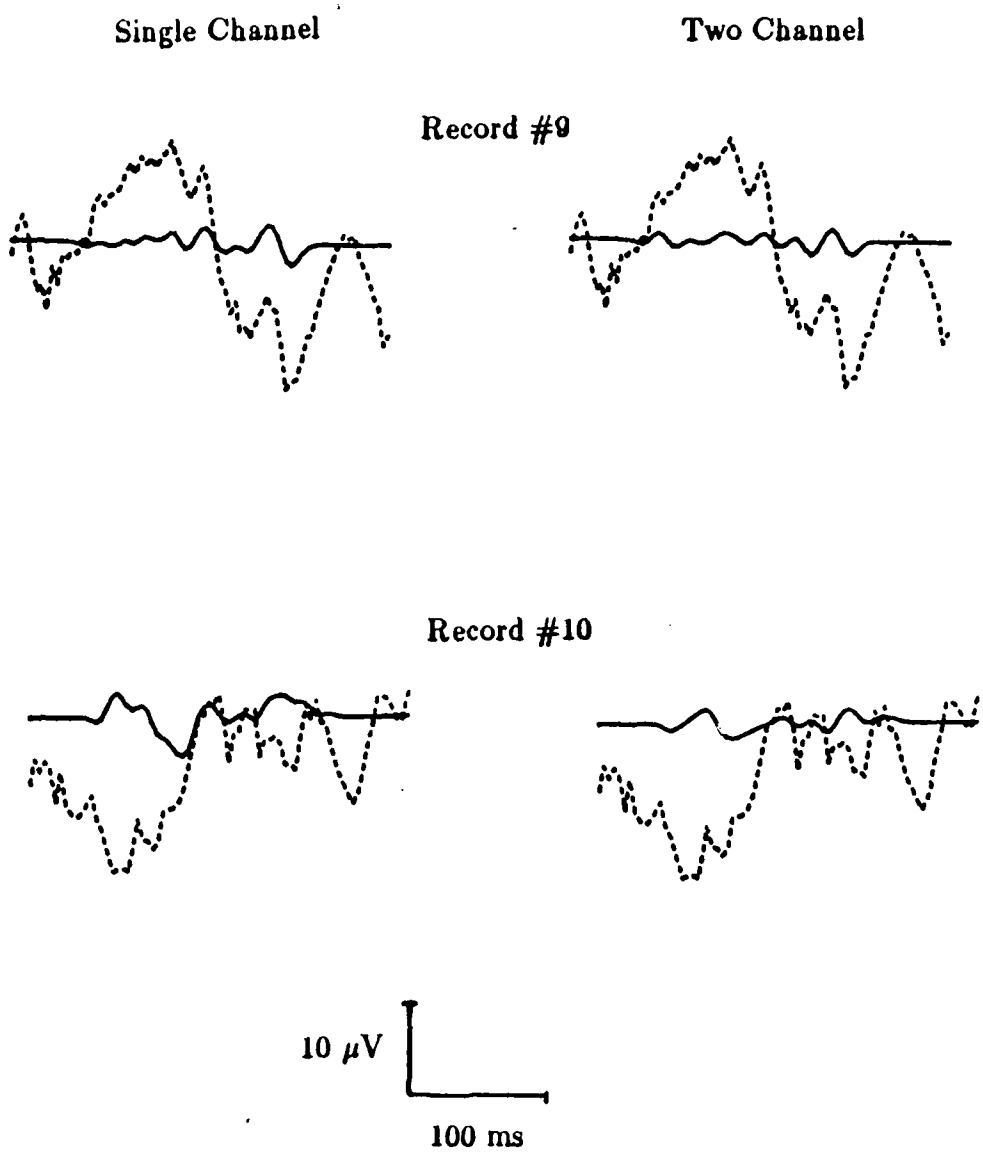


Figure 4-15 Comparison of Single and Two Channel Filter Output Using Prestimulus Electroencephalogram as Input (Subject #5). Pz Input (dash) and Filter Output (solid).

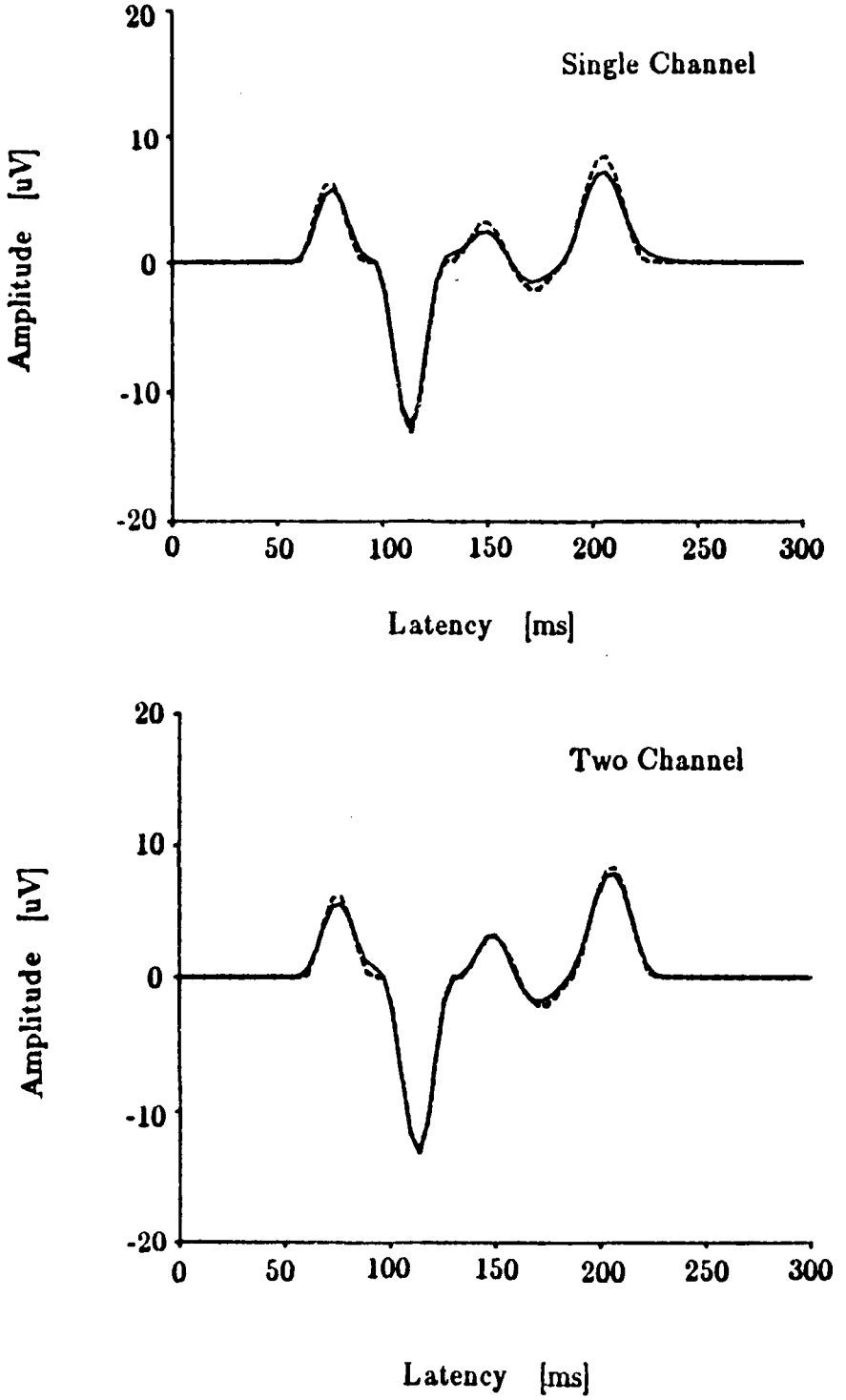


Figure 4-16 Comparison of Single and Two Channel Filter Outputs for Signal Only at Input (Subject #5 Signal Model). Signal Model (dash) and Filter Output (solid).

obtainablee using extra channels.

In order to better understand the importance of the MNLS criterion, a second two channel filter was designed using more eigenvalues than necessary to compute the pseudoinverse and solve the filter normal equations. The original l_{opt} truncated pseudoinverse filter was designated the LOPT filter and the non-l_{opt} truncated pseudoinverse filter was designated the NLOPT filter. The same scalp-recorded data records (Subject #5) used to design the LOPT filter were used to design the NLOPT filter. Table 4-8 summarizes the performances of the LOPT and NLOPT two channel filters. The key result is that even though the theoretical MSE is reduced by keeping 132 eigenvalues (instead of the 112 suggested by the MNLS criterion), the MNLS error is not. It is actually very large (since theoretically it cannot be greater than 1.0) suggesting that the filter output is unstable. Figure 4-17 depicts several examples of the NLOPT filter output along with the LOPT filter output for the identical input. It is obvious that the NLOPT filter output is unstable and that it is not minimizing the MSE. When designing *a posteriori* filters, therefore, it is necessary to use the MNLS criterion to account for possible instabilities in the filter due to the differences in the modeled cross-correlation matrices and those of the actual input process.

Table 4-8 Performance Comparisons Using LOPT and NLOPT Filters. LOPT Filter Retained 112 Eigenvalues to Compute Pseudoinverse and NLOPT Filter Retained 132. Abbreviations Used are MSE (Mean-Square Error), MNLSE (Minimum Norm Least Square Error), BF (Bias Factor), NRF (Noise Reduction Factor), SNR_I (Input Signal-to-Noise Ratio), and SNR_O (Output Signal-to-Noise Ratio). MSE and BF are Normalized.

Criterion	NLOPT	LOPT
Theoretical MSE	0.028	0.106
MNLSE	2.485	0.147
BF	0.022	0.048
SNR_I (dB)	-7.756	-7.756
NRF (dB)	-30.878	-20.109
SNR_O (dB)	22.967	11.576

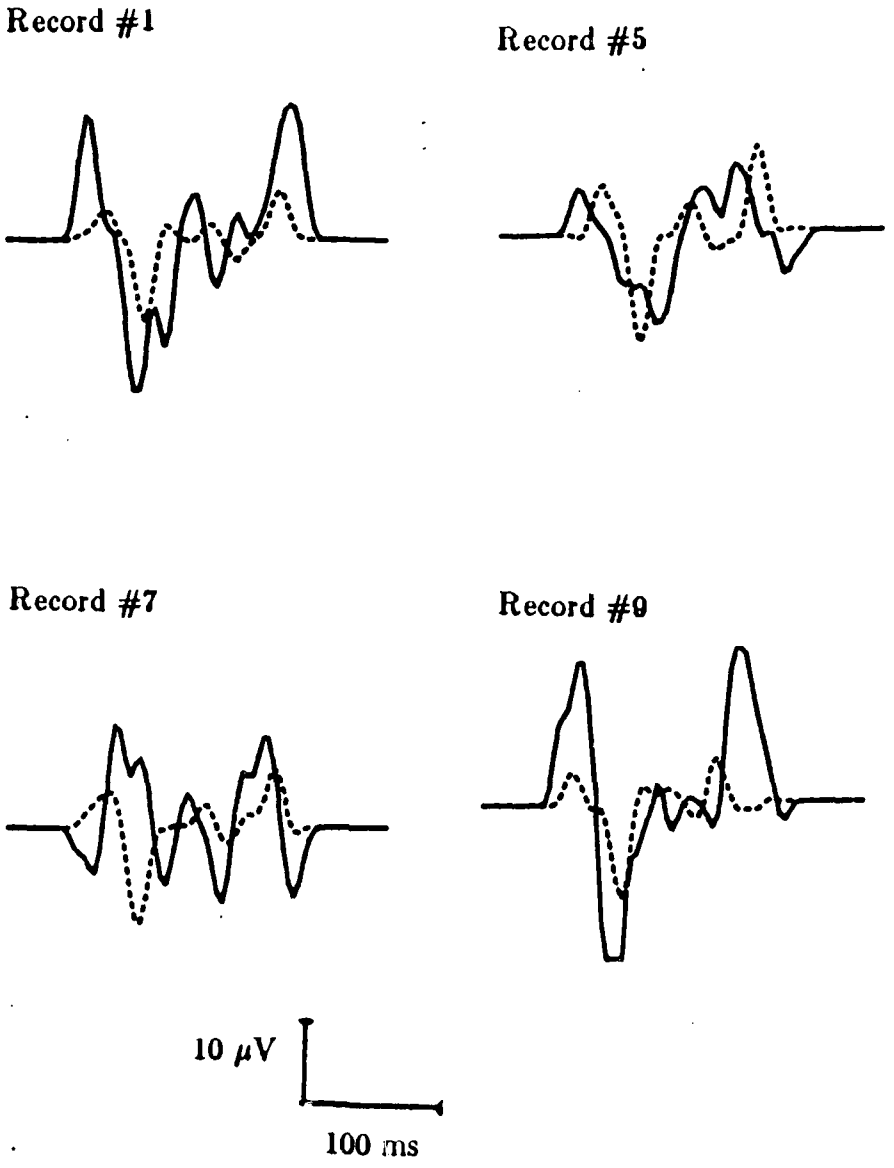


Figure 4-17 Two Channel LOPT Filter Output (dash) and Two Channel NLOPT Filter Output (solid). LOPT Filter Retained 112 Eigenvalues to Compute Pseudoinverse and NLOPT Filter Retained 132.

5. MATHEMATICAL MODELING OF THE VEP

5.1. Introduction

The object of this study was to mathematically model the ERP recorded at the surface of the cortex. These waves describe the net result of the electrical activity generated by the simultaneous activation of a large number of thalamocortical axons (Mountcastle 1974). In some recoding situations, a very precise correlation has been found between the ERP recorded with microelectrodes within the cortex and the surface recorded ERP (Creutzfeldt 1966 and Cooper 1965). Thus since both types of recordings are generated by the same events, they should be mathematically related. Yet, a one to one relationship between the potentials generated by a single neuron or section of a neuron and the recorded surface potential does not exist. The reasons for this are that the surface electrodes are large enough to record the activity of many cells and that the neuronal membrane potentials do not vary in a synchronous way. Nevertheless, representing the ERP as the addition of a set of components or basis functions which might be similar to these neuronal discharges will aid in the development of effective signal processing algorithms for extracting the ERP.

5.2. The Model

The model proposed is given by Equation (5-1)

$$\hat{x}(t) = \sum_{k=1}^{n} a_k s(t - t_k) \qquad (5\text{-}1)$$

where the coefficients $\{a_k\}$ are statistically independent identically distributed(i.i.d.) random variables as are the $\{ t_k \}$. The function of time $\hat{x}(t)$ is the approximation to the measured surface potential and $s(t)$ is an elementary waveform capable of representing a component in the ERP. This representation is useful because it provides a model for analyzing experimental results and for carrying out simulation studies. The average waveform is given by the expected value of $\hat{x}(t)$ which can be expressed as

$$\hat{x}(t) = \sum_{k=1}^{n} E[a_k]E[s(t-t_k)] \tag{5-2}$$

$$= \sum_{k=1}^{K} \bar{a}_k \int_{-\infty}^{\infty} s(t-t_k)p(t_k)\,dt_k$$

$$= \sum_{k=1}^{K} \bar{a}_k \tilde{s}(t-t_k)$$

where $p(t_k)$ is the probability density function of t_k, $\bar{a}_k$ is the mean of a_k, $\bar{t}_k$ is the mean of t_k, and $\tilde{s}$ is the convolution of the component waveform with the probability density function.

Two models similar to the above were used by Moser (Moser 1980). In one of his models the evoked potential was depicted as the sum of gaussian shaped wavelets multiplied by an amplitude factor as shown in Equation (5-3)

$$\hat{x}(t) = \sum_{k=1}^{n} \alpha_k \exp[-\frac{(t-t_j)^2}{2\sigma^2}] \tag{5-3}$$

where the $\hat{x}(t)$ represented the measured surface potential, the α_k were the amplitude factors for each of the k-th components, σ was a measure of the width of the wavelets and t_j was a random variable representing the time of occurrence of the wavelet's peak. An alternative model (Yu 1983) was based on the same principle but the wavelet shape used was that of a raised cosine pulse.

5.3. Statistical Analyses

In most of the models mentioned as well as in other applications aimed toward the improvement of waveform estimation of event-related potentials (McGillem 1985) the assumption has been made the the latencies of each component behave as if they were normally distributed. This assumption is supported by the histograms of the latencies of individual peaks detected during computation of the LCA which indicate that the gaussian density function is a reasonable approximation for their distribution. Also, it seems likely that the latency of each component corresponds to the instant in time at which the greatest number of neurons fired. The large number of contributing neurons also suggests a gaussian behavior.

In order to strengthen the supposition, some preliminary nonparametric statistical tests have been performed on evoked potential data. The data were

obtained from subjects who sat in a dark and quiet room viewing a cathode-ray-tube screen(Grass System) which subtended a 11.5 by 8.5 visual angle. Beckman silver-silver chloride electrodes were attached on their heads with conductive paste at Cz, Pz, Oz, and In, according to the 10-20 system (Jasper 1958). All the electrodes were referenced to linked mastoids and the forehead was used as ground. Interelectrode impedance was kept below 5 kohms . The stimulus consisted of upper and lower checkerboard patterns of check size approximately 18'. The stimulus duration was 1000 msec. with interstimulus interval varying between 3 to 5 sec. The average luminance was kept constant throughout the experiment at 6.0 foot Lamberts. Grass 7p511 EEG amplifiers with a low frequency cutoff of 0.1 Hz and high frequency cut off of 100 Hz were used. Analog-to-digital conversion was performed at a rate of 250 samples per second with an A/D converter having 12 bit precision. The data was later searched for artifacts due to eye movement. If the eye channel signal amplitude changed more than 50 μv in 100ms the record was rejected.

The tests were only executed on the lower checkerboard pattern data recorded at Oz and Pz. The first step was to acquire from the LCA program the latency and amplitude values for each component that was a member of the sample population. The statistics for each peak in the signal are given in Tables 5-21 & 5-22 and only those peaks detected more than 30% of the time were later analyzed.

Four different nonparametric statistical methods were used. These are the run test, the chi-square goodness of fit, the Kolmogorov-Smirnov one sample test and the Kolmogorov-Smirnov two sample test. These tests were chosen mainly because they are distribution-free techniques and do not depend on the characteristics of the population from which the samples are acquired. In fact there are only two main requirements, one being that the sample variates be continuously distributed and the other is for the observations to be drawn randomly and independently of the outcome of previous draws.

The first requirement was assumed true while the second was tested using the run test. The hypothesis of independence was tested at the .05 level of significance and all the latencies and amplitudes variates for each component passed the test. The results are presented in Tables 5-1 through 5-4 where the number of runs is given under the label r and the columns to its left and right represent the limiting values of r under which the hypothesis of independence is accepted.

The skew and kurtosis of the variables were also measured. The skew describes the degree of asymmetry of a density function and it is zero for a

Table 5-1 Results of Run Test on Latencies (Oz)

Peak #	N	α	$r_{n/2,1-\alpha/2}$	r	$r_{n/2,\alpha/2}$
p1	36	.05	12	18	25
p2	77	.05	29.8	33	48.2
p3	95	.05	38	45	58
n1	48	.05	17.2	26	31.8
n2	97	.05	38.8	48.5	59.2
n3	45	.05	16	24	30

Table 5-2 Results of Run Test on Latencies (Pz)

Peak #	N	α	$r_{n/2,1-\alpha/2}$	r	$r_{n/2,\alpha/2}$
p1	79	.05	30.6	35	49.4
p2	50	.05	18	27	33
p3	80	.05	31	34	50
n1	32	.05	11	17	22
n2	99	.05	39.6	53	60
n4	30	.05	18.44	26	33.7

Table 5-3 Results of Run Test on Amplitudes (Oz)

Peak #	N	α	$r_{n/2,1-\alpha/2}$	r	$r_{n/2,\alpha/2}$
p1	36	.05	12	19	25
p2	77	.05	29.8	43	48.2
p3	95	.05	38	49	58
n1	48	.05	17.2	21	31.8
n2	97	.05	38.8	53	59.2
n4	45	.05	16	19	30

Table 5-4 Results of Run Test on Amplitudes (Oz)

Peak #	N	α	$r_{n/2,1-\alpha/2}$	r	$r_{n/2,\alpha/2}$
p1	79	.05	30.6	39	49.4
p2	50	.05	18	31	33
p3	80	.05	31	41	50
n1	32	.05	11	21	22
n2	99	.05	39.6	47	60
n4	30	.05	18.44	27	33.7

normal distribution. The kurtosis describes the extend to which a frequency distribution is concentrated about the mean and it is equal to three for gaussian distributions. The skew is defined as the ratio of the third moment to the cube of the standard deviation and the kurtosis is defined as the ratio of the fourth moment to the square of the second central moment.

The meaning of the values obtained for the skew is difficult to assess since they do not follow any distinct pattern. Yet, it should be pointed out that if the figures in the tables are rounded off to the nearest integer, in only one case of the amplitude measurements does the skew deviate from zero, while in the case of the latency measurements, it deviates on three occasions. It therefore seems that the amplitude distributions tend to be more symmetric than the latency distributions.

The significance of the figures derived for the kurtosis is also hard to interpret. Nevertheless, it can be said that in both the amplitude and latency cases most of the distributions are platykurtic(kurtosis$<$3) with a few leptokurtic ones(kurtosis$>$3). Once again, if the values are rounded off to the nearest integer, most of the distributions assume mesokurtic characteristics.

The next test performed was the Chi-Square goodness-of-fit which is one of the best known and more used distribution-free procedures for data evaluation. The scope of its utility is limited because of requirements that can only be fulfilled when the sample size is infinite. For instance, the assumption that the chi-square distribution supplies a good approximation for the distribution of the test statistic is true if, the number of observations tends to infinity. Likewise some consideration must be given to the number of intervals that are used, since it may affect the resultant probabilities (Williams 1950). In any event the test is simple and a subroutine such as the one existing in IMSL simplifies its usage.

Tables 5-9 through 5-12 present the results of using the chi-square test. The first column identifies the peak, the second column indicates the number of cells into which the observations were distributed, The third indicates the degrees of freedom, the fourth the computed chi-square statistics and the last provides the probability of the null hypothesis being true.

The figures in the tables indicate that the amplitude variates satisfy the null hypothesis, but not the latencies. Nevertheless, the number of observations in most of the cases is less than 100, thus the results can not be totally accepted.

The next test used was the Kolmogorov-Smirnov one sample test which is exact for small sample sizes. There is some controversy over which test is more

Table 5-5 Skew and Kurtosis for Latencies (Oz)

peak	p1	p2	p3	n1	n2	n3
skew	0.29	0.12	-0.57	-0.13	0.04	-0.08
kurtosis	1.71	3.18	2.90	1.70	2.78	2.17

Table 5-6 Skew and Kurtosis for Latencies (Pz)

peak	p1	p2	p3	n1	n2	n4
skew	-0.10	-0.63	0.17	-0.19	0.02	-0.66
kurtosis	3.18	2.63	2.59	1.69	3.81	2.50

Table 5-7 Skew and Kurtosis for Amplitudes (Oz)

peak	p1	p2	p3	n1	n2	n3
skew	-0.36	0.32	0.01	-0.15	-0.05	-0.04
kurtosis	2.43	2.96	2.77	2.58	2.75	2.61

Table 5-8 Skew and Kurtosis for Amplitudes (Pz)

peak	p1	p2	p3	n1	n2	n3
skew	0.16	-0.06	-0.23	0.75	-0.17	0.02
kurtosis	3.96	2.22	3.28	4.19	2.55	2.98

Table 5-9 Results of Chi-Square Test on Latencies (Oz)

Peak #	# of cells	df	cs	P
p1	8	5	24.89	0.00
p2	8	5	32.61	0.00
p3	8	5	12.54	0.03
n1	8	5	24.67	0.00
n2	8	5	39.16	0.00
n3	8	5	16.69	0.01

Table 5-10 Results of Chi-Square Test on Latencies (Pz)

Peak #	# of cells	df	cs	P
p1	8	5	62.06	0.00
p2	8	5	12.40	0.03
p3	8	5	6.00	0.31
n1	8	5	30.50	0.00
n2	8	5	53.48	0.00
n4	8	5	29.16	0.00

Table 5-11 Results of Chi-Square Test on Amplitudes (Oz)

Peak #	# of cells	df	cs	P
p1	8	5	3.11	0.68
p2	8	5	7.05	0.22
p3	8	5	1.76	0.88
n1	8	5	11.00	0.05
n2	8	5	00.73	0.98
n3	8	5	1.40	0.92

Table 5-12 Results of Chi-Square Test on Amplitudes (Pz)

Peak #	# of cells	df	cs	P
p1	8	5	5.56	0.35
p2	8	5	4.72	0.45
p3	8	5	8.00	0.16
n1	8	5	5.50	0.36
n2	8	5	8.56	0.13
n4	8	5	3.43	0.63

powerful but, in general, the Kolmogorov-Smirnov is considered to be more efficient. Further comparisons of the tests are presented by Slakter (1965).

Tables 5-13 through 5-16 present the outcome resulting from the IMSL subroutine NSK1. The first column gives the peaks, the second column presents the Kolmogorov statistics, and the third the probability for accepting the hypothesis of equality. The results substantiate those obtained by the chi-square test. In other words, the amplitude variates are normally distributed random variables while the latency variates are not.

Another supposition for the model that can be tested is whether the amplitude variables and latency variables are respectively, identically distributed random variables. This is done by applying the Kolmogorov-Smirnov two-sample test. This test resembles the one-sample test and the final outcomes are given in Tables 5-17 through 5-20. The results indicates that this condition is only true for the amplitude case.

These results are surprising but they can not be taken as final since more data is needed to reach a conclusive decision. However they have not provided a satisfactory response to the initial question about the type of distribution these variables have. Current research is aimed toward understanding the effects of removing the best-fit linear trend on the results of the statistical test, and the effects of the technique that it is used to define the latency range of each peak. The studies consist of generating and testing data under the characteristics of the model specified by Equation (5-1). The nonparametric tests are run on the data before and after removing the best-fit linear trend and so far the indications point toward the fact that the results are not affected by this. A similar test will be performed to review the effects of the range finding technique.

5.4. Simulation Tests

Up to now nothing has been said about the major difference between this model and those used before. This is the acknowledgment that certain of the observed peaks are just "valleys" between contiguous components of the same polarity and not an isolated component. Determining which peaks are components and which are artifacts is not a simple task and can not be done by just examining the latency corrected average or average of the signal. An investigation of the shapes of the signal peaks and the correlation between their latencies and amplitudes must be performed.

It was suggested that the movement of a "valley" must be highly correlated with the movement of the two positive peaks which give origin to it

Table 5-13 Results of Kolmogorov-Smirnov I Test on Latencies (Oz)

Peak #	D	P
p1	0.33	0.0009
p2	0.32	0.00000012
p3	0.21	0.000325
n1	0.34	0.000039
n2	0.30	0.00
n3	0.29	0.00084

Table 5-14 Results of Kolmogorov-Smirnov I Test on Latencies (Pz)

Peak #	D	P
p1	0.17	0.0217
p2	0.15	0.21
p3	9.54	0.46
n1	0.22	0.0866
n2	0.13	0.0627
n4	0.16	0.17

Table 5-15 Results of Kolmogorov-Smirnov I Test on Amplitudes (Oz)

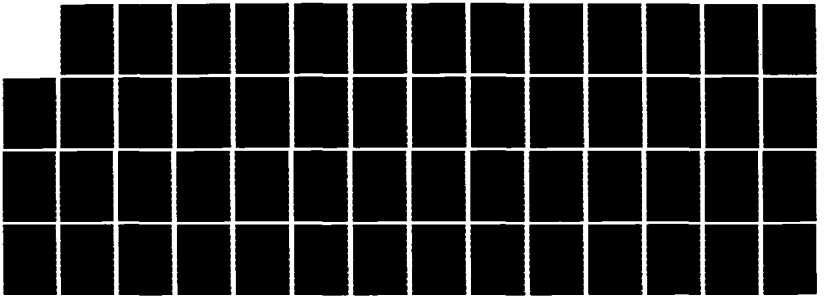
Peak #	D	P
p1	0.0850	0.96
p2	0.0837	0.65
p3	0.0520	0.96
n1	0.0714	0.97
n2	0.0399	0.998
n3	0.0637	0.99

Table 5-16 Results of Kolmogorov-Smirnov I Test on Amplitudes (Pz)

Peak #	D	P
p1	0.0814	0.67
p2	0.0756	0.94
p3	0.0839	0.63
n1	0.0985	0.92
n2	0.0726	0.67
n4	0.0782	0.91

AD-A170 125 DETECTION ESTIMATION AND MULTIDIMENSIONAL PROCESSING OF SINGLE EVOKED POT..(U) PURDUE UNIV LAFAYETTE IN EEG SIGNAL PROCESSING LAB J I AUNON ET AL. SEP 85 2/2

UNCLASSIFIED AFOSR-TR-86-0367 F49620-83-K-0031 F/G 6/16 NL



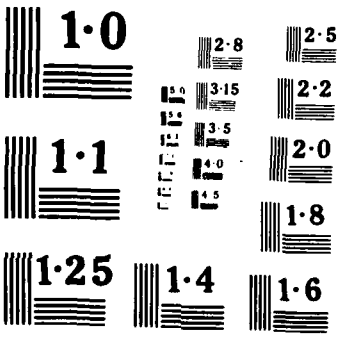

1·0
1·1
1·25
1·4
1·6
1·8
2·0
2·2
2·5
2·8
3·15
3·5
4·0
4·5

Table 5-17 Results of Kolmogorov-Smirnov II Test on Latencies (Oz)

Peak #	D	P
p1-p2	0.21	0.17
p1-p3	0.25	0.0571
p1-n1	0.27	0.0732
p1-n2	0.23	0.0974
p1-n3	0.22	0.22
p2-p3	0.22	0.0235
p2-n1	0.23	0.0893
p2-n2	0.24	0.017
p2-n3	0.19	0.19
p3-n1	0.22	0.0837
p3-n2	0.22	0.0143
p3-n3	0.22	0.0939
n1-n2	0.23	0.0508
n1-n3	0.19	0.31
n2-n3	0.20	0.15

Table 5-18 Results of Kolmogorov-Smirnov II Test on Latencies (Pz)

Peak #	D	P
p1-p2	0.25	0.0388
p1-p3	0.25	0.0125
p1-n1	0.27	0.0616
p1-n2	0.29	0.0812
p1-n3	0.20	0.13
p2-p3	0.22	0.0815
p2-n1	0.35	0.0131
p2-n2	0.22	0.0732
p2-n4	0.23	0.0959
p3-n1	0.32	0.0146
p3-n2	0.22	0.0253
p3-n4	0.25	0.0348
n1-n2	0.24	0.0909
n1-n4	0.41	0.0020
n2-n4	0.17	0.24

Table 5-19 Results of Kolmogorov-Smirnov II Test on Amplitudes (Oz)

Peak #	D	P
p1-p2	0.14	0.63
p1-p3	0.12	0.77
p1-n1	0.13	0.78
p1-n2	0.0767	0.99
p1-n3	0.13	0.78
p2-p3	0.0801	0.91
p2-n1	0.0906	0.94
p2-n2	0.0952	0.77
p2-n3	0.0987	0.90
p3-n1	0.0798	0.97
p3-n2	0.0598	0.99
p3-n3	0.0702	0.99
n1-n2	0.0872	0.94
n1-n3	0.0833	0.98
n2-n3	0.0733	0.99

Table 5-20 Results of Kolmogorov-Smirnov II Test on Amplitudes (Pz)

Peak #	D	P
p1-p2	0.11	0.77
p1-p3	0.0957	0.79
p1-n1	0.13	0.76
p1-n2	0.14	0.30
p1-n4	0.10	0.86
p2-p3	0.11	0.76
p2-n1	0.12	0.86
p2-n2	0.12	0.68
p2-n4	0.12	0.75
p3-n1	0.16	0.56
p3-n2	0.13	0.39
p3-n4	0.11	0.73
n1-n2	0.0972	0.95
n1-n4	0.14	0.75
n2-n4	0.0772	0.97

as it can be seen in the simple case illustrated below. In this illustration the components are given by Equation (5-5)

$$s_1(t)=a_1\exp[-\frac{(t-m_1)^2}{b_1^2}] \qquad (5\text{-}5)$$

$$s_2(t)=a_2\exp[-\frac{(t-m_2)^2}{b_2^2}]$$

For the purpose of illustrating the previous statement the variables a_1 and a_2 are considered to be deterministic and of equal value. Also the variables b_1 and b_2 were set to one. The final waveform is given by the sum of these two wavelets as shown in Equation (5-6)

$$x(t)=s_1(t)+s_2(t)=a_1\exp[-\frac{(t-m_1)^2}{b_1^2}]+a_2\exp[-\frac{(t-m_1)^2}{b_1^2}] \qquad (5\text{-}6)$$

Setting the derivative to zero and solving for t provides the location of the false peak which is given in Equation (5-7)

$$t=\frac{a_1+a_2}{2} \qquad (5\text{-}7)$$

which is the average value of the latencies of the contiguous peaks. In the case a negative component exist, the correlation between its latency jitter and that of its adjacent peaks must be low although it will probably depend on the amplitudes of these nearby components.

In order to corroborate this statement and search for ways of differentiating between false and real components data following the properties of Equation (5-4) was generated. Six sets of signals were produced. The first three consisted of two positive gaussian shaped components with amplitudes varying from one to two units. Their mean latency was taken to be 78.03 and 176.46 ms which are the corresponding means for the second and third positive peaks in the LCA as can be seen from Table 5-21. The standard deviation for the latencies was taken from the same table and they are 7.04 and 14.49 ms, respectively. The width of the peaks was set to 22 ms, this value might a be little high when compared to the width of the components of the evoked potential but it assures the interaction of the components simulated. Figure 5-1 shows nine out of the one hundred signals generated for each set. Columns one ,three and five represent the two positive components before they are added together while columns two,four and six are the result of their summation. The

Table 5-21 LCA Results for Subject #5 (Oz)

Positive Peaks							
peak #	range	mean	st.d.	pct.	max.	min.	amp.st.d.
1	6-10	26.67	5.66	36.00	4.25	-1.28	6.03
2	17-26	78.03	7.04	77.00	5.28	-4.57	7.32
3	34-51	176.46	14.49	95.00	5.32	0.99	7.26
Negative Peaks							
1	12-18	56.75	7.41	48.00	3.13	-2.90	7.07
2	24-35	113.90	7.44	97.00	-1.71	-9.14	6.77
3	53-59	220.98	7.44	45.00	2.27	-2.21	7.62
4	63-67	256.43	6.38	28.00	1.56	-2.42	5.81
5	69-71	276.87	3.18	23.00	0.44	-4.26	9.25

Table 5-22 LCA Results for Subject #5 (Pz)

Positive Peaks							
peak #	range	mean	st.d.	pct.	max.	min.	amp.st.d.
1	16-25	75.65	5.74	79.00	6.23	-5.93	11.39
2	32-41	146.40	9.21	50.00	2.88	-3.48	9.85
3	46-58	202.75	11.52	80.00	8.96	3.01	13.01
Negative Peaks							
1	12-17	55.00	5.56	32.00	3.18	-3.75	9.12
2	23-34	111.76	6.78	99.00	-3.09	-13.11	15.80
3	41-46	168.93	5.82	30.00	1.56	-2.87	10.30
4	65-71	271.92	7.57	51.00	1.54	-3.37	8.28
5	69-71	276.87	3.18	23.00	0.44	-4.26	9.25

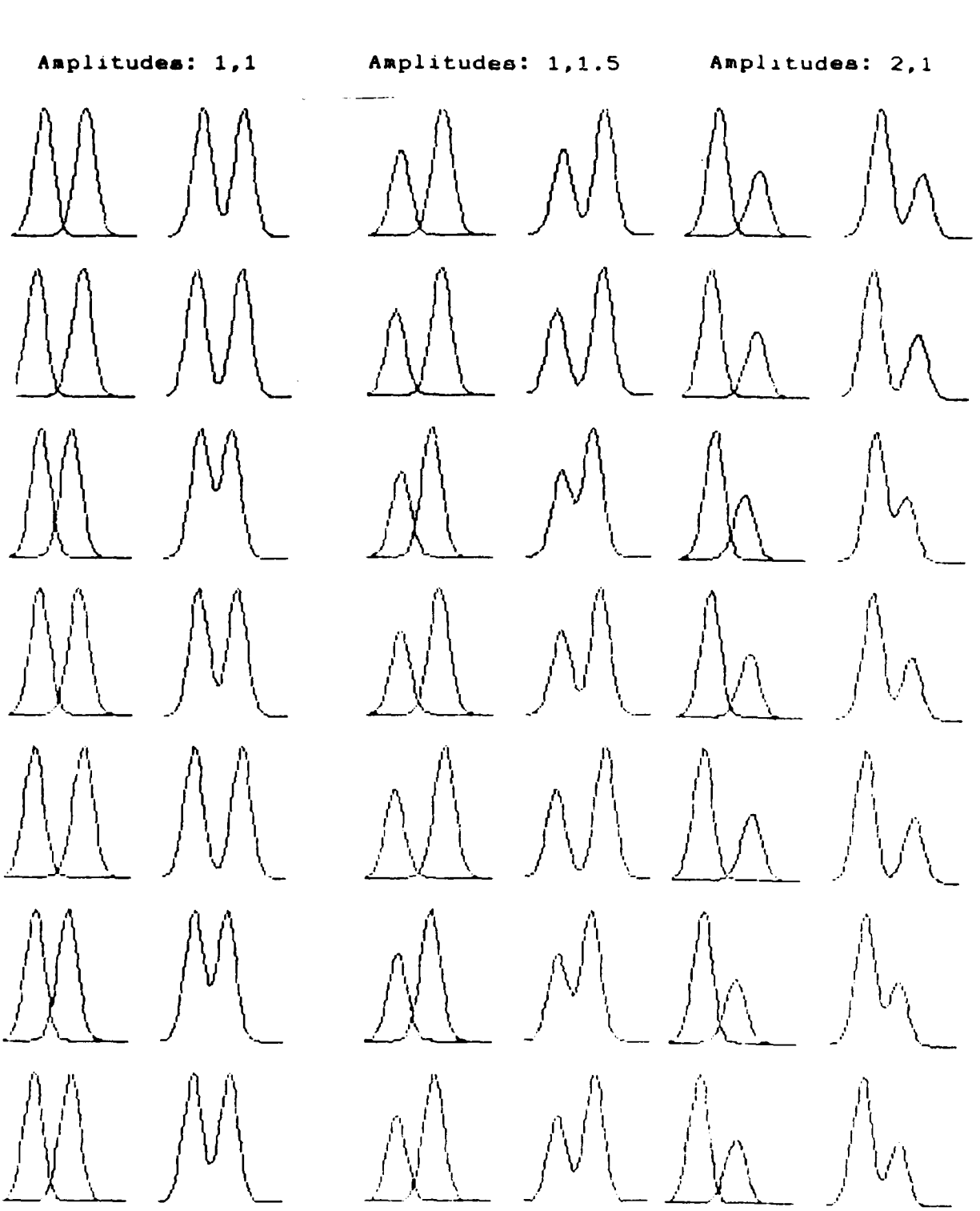


Figure 5-1 Signals Formed By Adding Two Positive Peaks

vertical scale varies from 0.0 to 2.0 while the horizontal scale goes from 0.0 to 3.0 ms. The other three sets of signals are similarly illustrated in Figure 5-2 with the exception that the vertical scale varies from -2.0 to 2.0. These sets of signals were generated by adding two positive peaks with the same characteristics as those described above and a negative peak. The latency of the negative peak was assumed to be 113.90 ms which from Table 5-21 is seen to correspond to the second negative peak in the LCA. The standard deviation of its latency was taken to be 7.44 ms and its width was kept at 22 ms. Note that since the position of the baseline is known, it is easy to identify which signals contained false components and which have a negative signal. In evoked potentials, however, the actual position of the baseline is not known and distinguishing between "valleys" and negative peaks is not as easy.

Once the sets of signals were generated, each of the 100 signals in a set was searched for the position of its peaks. The process consisted of successively computing the difference between three consecutive points and storing their location and values. A quadratic polynomial was then fitted to these three points obtaining in this manner the position of the maxima and minima. In order to detect the amount each peak moved from its presupposed location(the mean), the average latency of each component was subtracted from the measurement taken. The average movement of the two positive peaks was computed and compared to the movement of the valley or negative peak between them.

Table 5-23 presents the correlation coefficients obtained between the average movement of the two positive peaks and the valley or negative component in between. Note that when the two positive components have the same amplitude and, no negative peak exists, the correlation coefficients had a value of one. In fact in all the cases were the signal consisted of the sum of two positive peaks, the correlation was above .9. When the signal contained a negative component, the correlation was much lower with one exception occurring when the amplitude of the positive peaks was twice that of the negative one.

Figures 5-3 through 5-8 are plots of the negative peak latency shift versus the average latency shift of the positive peaks and are another way of illustrating the relationship between them. These figures can be thought of as being the sample distributions of two dimensional classes with different mean vectors and covariance matrices and can be used to estimate the probability of error in discriminating these two classes. The probability of error was computed using the algorithm developed by Fukunaga and Krile (1972) which

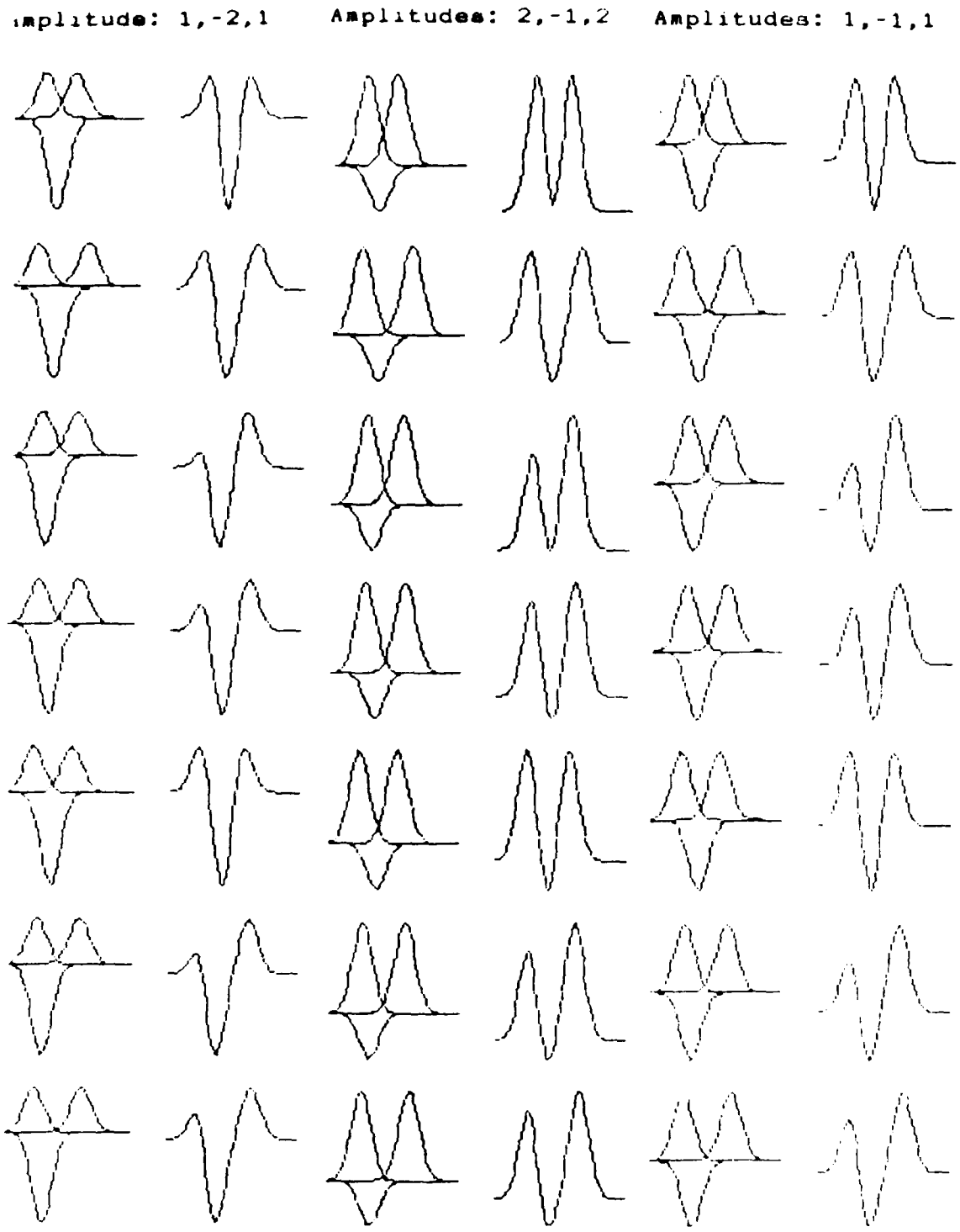


Figure 5-2 Signals Formed By Adding Two Positive Peaks and One Negative Peak

Table 5-23 Correlation Coefficients Between the Amplitudes of Actual and Artifactual Peaks

Signal: Sum of two positive peaks	
Amplitudes of peaks	Correlation
1,1	1.00
1,1.5	0.99
2,1	0.94
Signal: Sum of two positive and a negative peak	
1,-1,1	0.49
1,-2,1	0.27
2,-1,2	0.77

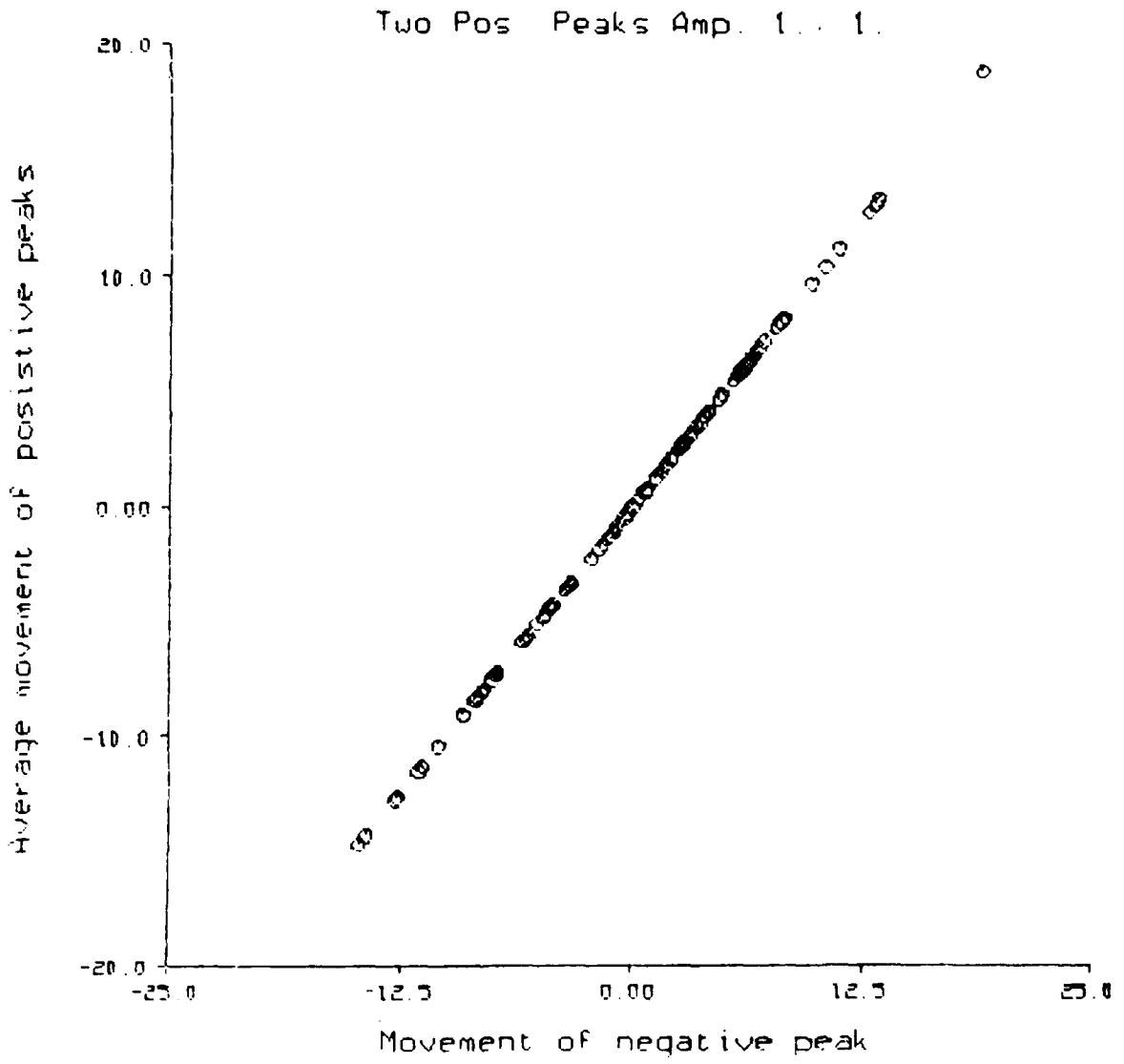


Figure 5-3 Average Movement of Positive Peaks vs. Movement of Negative Peak in Signals Formed By Adding Two Positive Peaks (Normalized Amplitudes 1. and 1.)

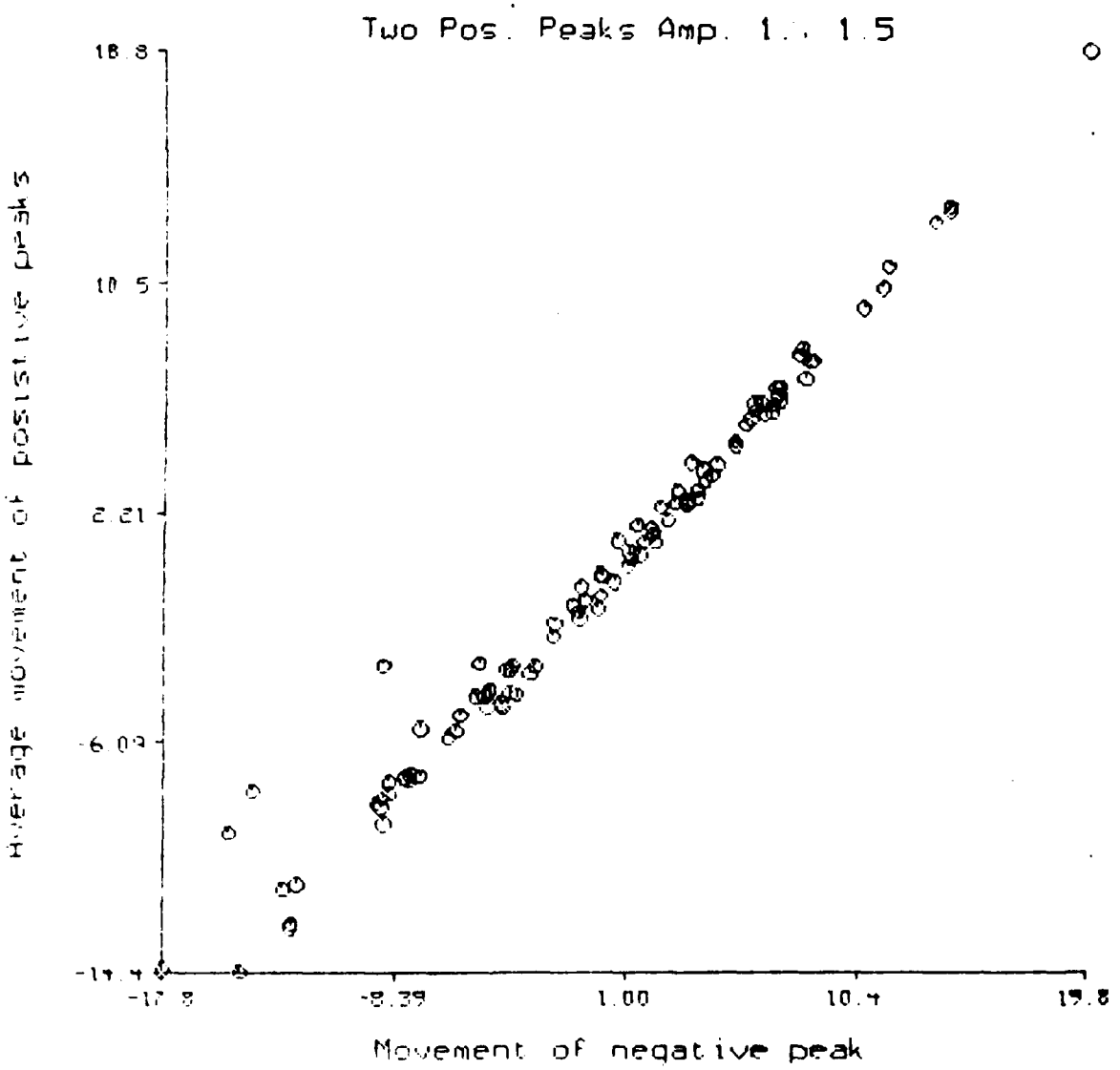


Figure 5-4 Average Movement of Positive Peaks vs. Movement of Negative Peak in Signals Formed By Adding Two Positive Peaks (Normalized Amplitudes 1. and 1.5)

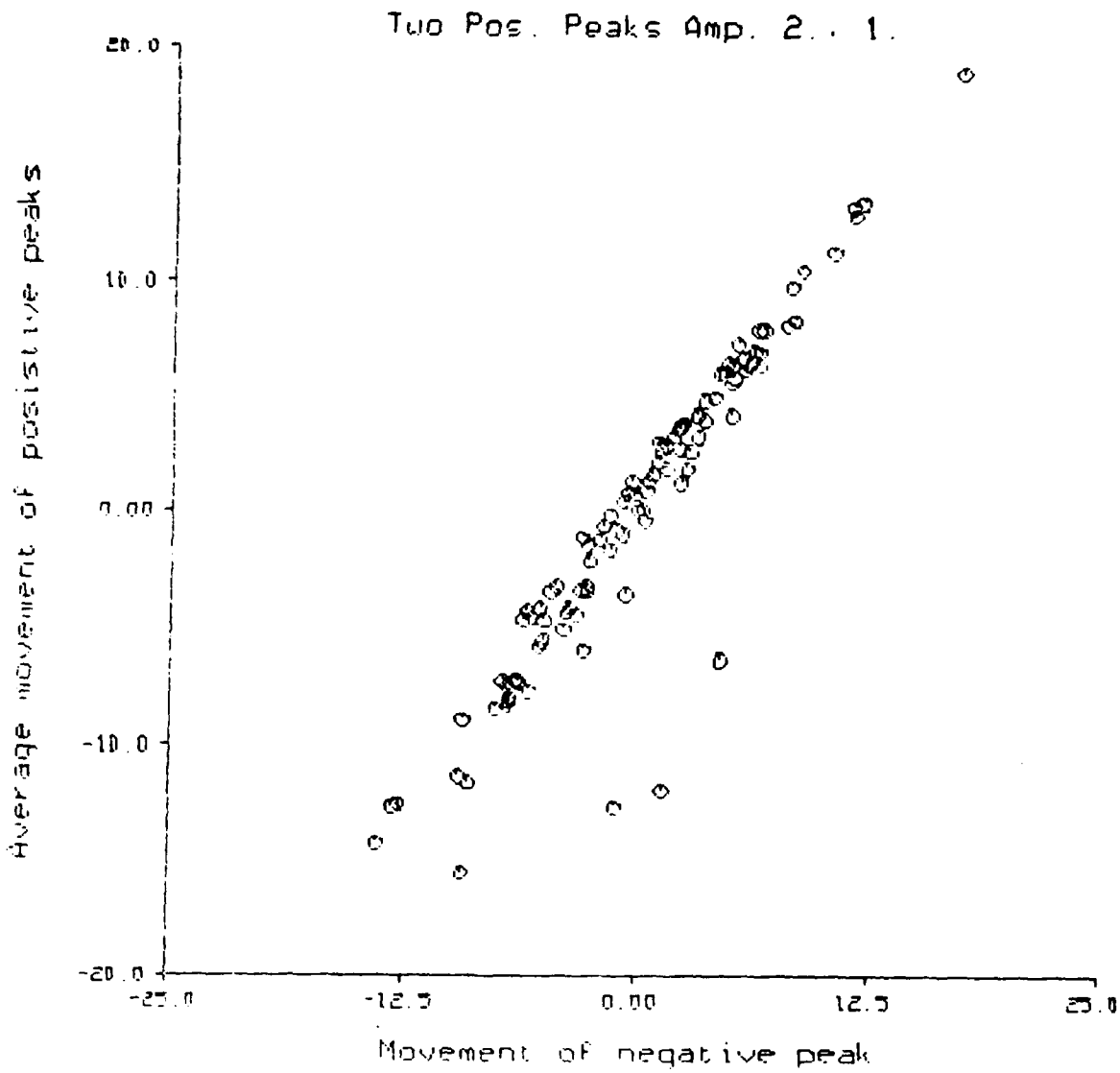


Figure 5-5 Average Movement of Positive Peaks vs. Movement of Negative Peak in Signals Formed By Adding Two Positive Peaks (Normalized Amplitudes 2. and 1.)

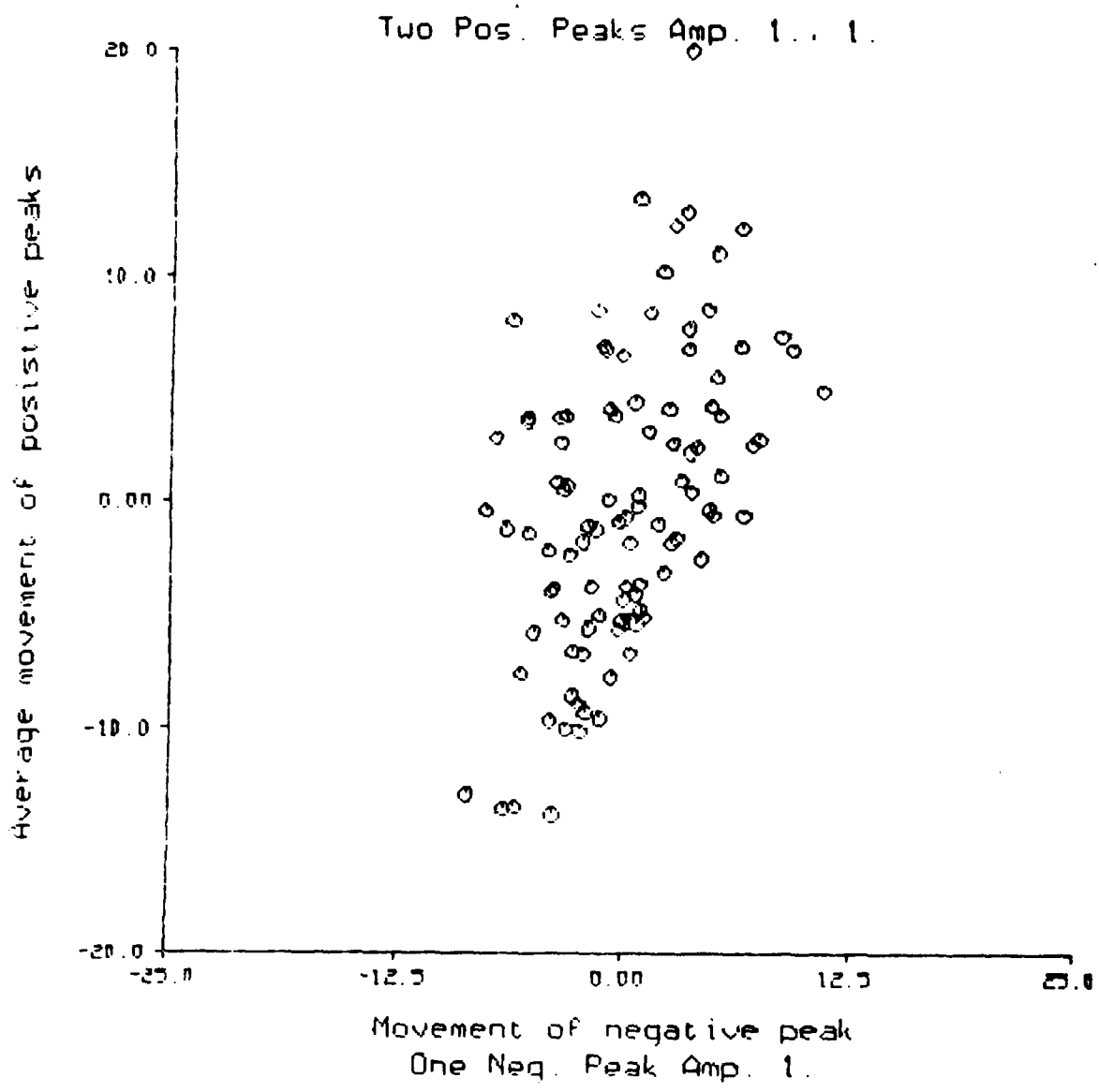


Figure 5-6 Average Movement of Positive Peaks vs. Movement of Negative Peak in Signals Formed By Adding Two Positive Peaks and One Negative Peak (Normalized Amplitudes 1., -1., and 1.)

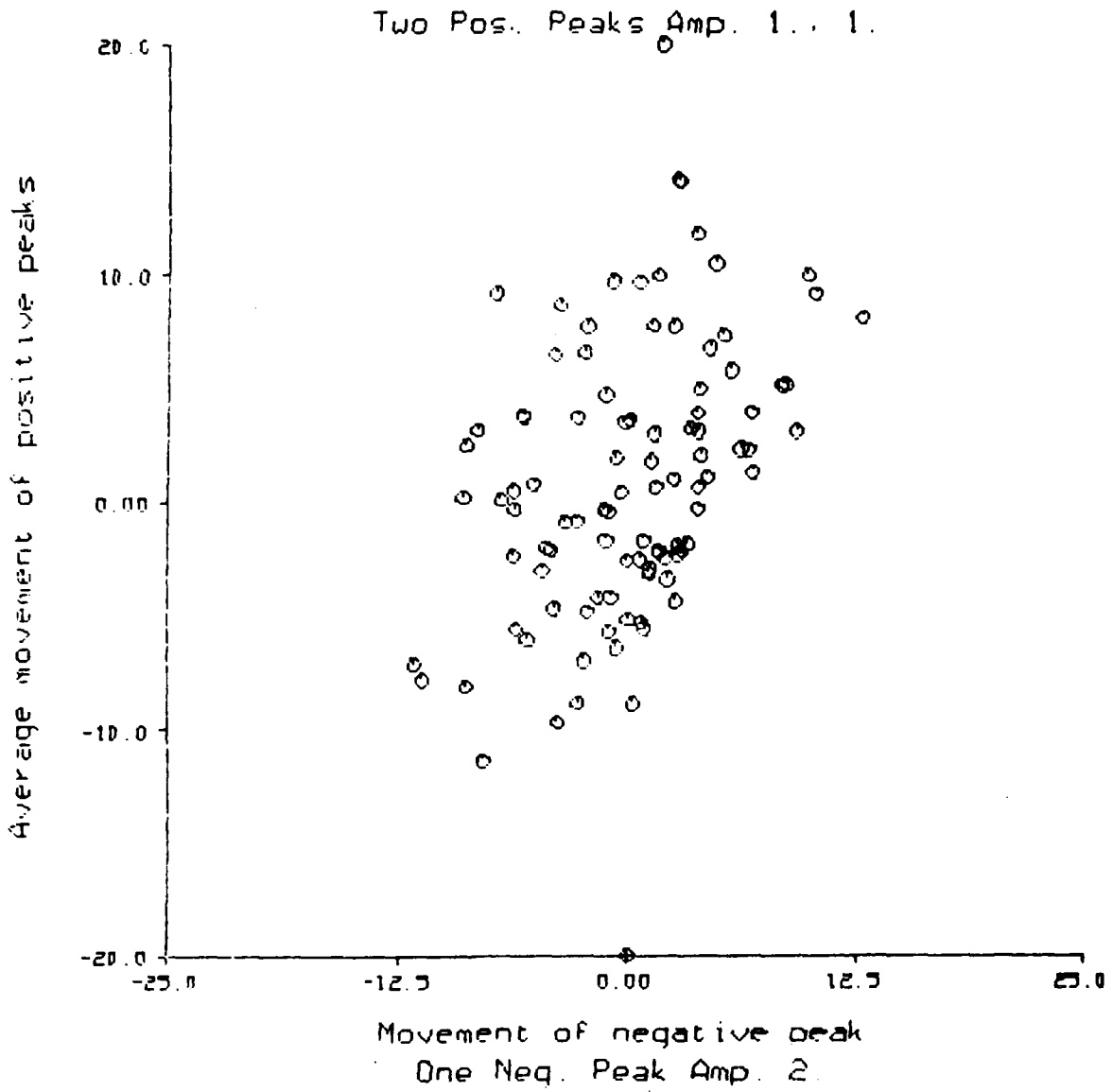


Figure 5-7 Average Movement of Positive Peaks vs. Movement of Negative Peak in Signals Formed By Adding Two Positive Peaks and One Negative Peak (Normalized Amplitudes 1., -2., and 1.)

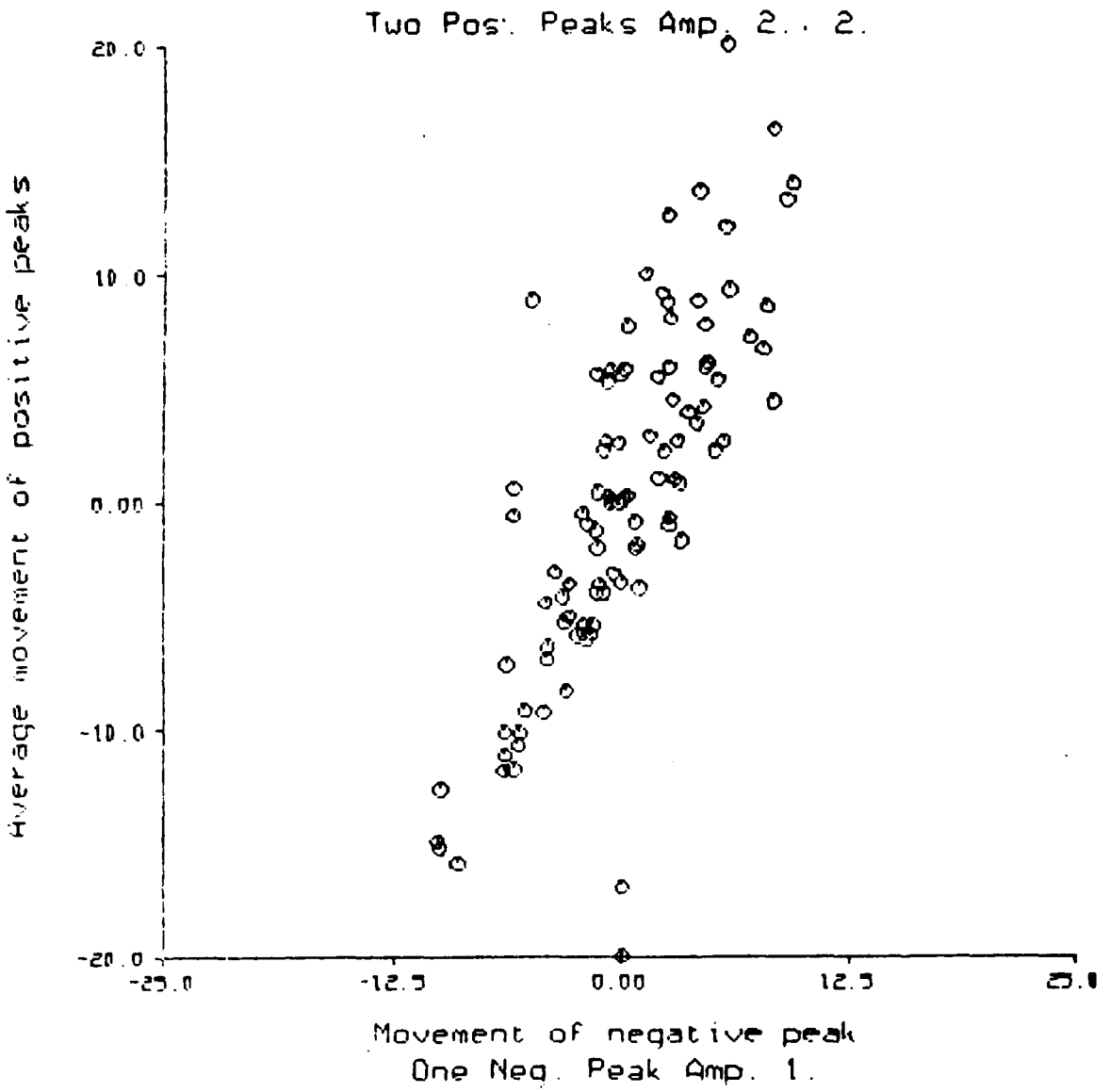


Figure 5-8 Average Movement of Positive Peaks vs. Movement of Negative Peak in Signals Formed By Adding Two Positive Peaks and One Negative Peak (Normalized Amplitudes 2., -1., and 2.)

converts multidimensional integration to one dimensional integration. In most cases, the resulting probability of error was 22% which indicates a classification accuracy of 78%. This result can probably be improved by using other features that might help in performing the differentiation between the two classes. One such feature might be found by studying the relationship between the amplitudes of the components and the effe s they might have on the amplitudes and latencies of the adjacent components.

5.5. Conclusions

Based on the preliminary studies that have been carried out, it appears likely that there is sufficient interaction between the latencies of artifactual components and the adjacent true components to make possible the identification of artifacts in a number of cases. A great deal of further study involving both simulation and measured data will be required before the most effective methods of making this determination are found.

6. DATA ACQUISITION SYSTEM IMPROVEMENTS

6.1. Modification

The EEG Signal Processing Laboratory has improved its data acquisition capabilities considerably with the addition of an IBM Personal Computer (PC). Figure 6-1 describes the current system. The IBM computer performs three important tasks: (1) experiment timing, (2) stimulus presentation, and (3) data collection. The PC includes a TECMAR Lab Master board with the following hardware:

(1) one 16 channel 12 bit analog-to-digital (A/D) converter with a maximum sampling rate of 40 kHz,

(2) two 12 bit digital-to-analog converters (D/A),

(3) five 16 bit counters, and

(4) three 8 bit parallel ports.

In addition, the PC has two 360 kbyte floppy disks for storing sampled data and a baud rate selectable serial port for transferring sampled data files to the PDP 11/45. The data files can then be transferred to the Engineering Computer Network (ECN) for signal processing on a Vax 11/780. The PC is also equipped with an 8087 math coprocessor which performs hardware multiply and divide operations at ten times the speed of the same software operations.

As depicted in Figure 6-1, subjects are seated in an IAC environmental chamber which is soundproofed and electromagnetically shielded. Scalp-electrode leads pass outside the chamber into a bank of Grass 7P511 EEG amplifiers and bandpass filters. The outputs of the EEG amplifiers are connected via coaxial cables to the input of the TECMAR A/D unit. Sampling

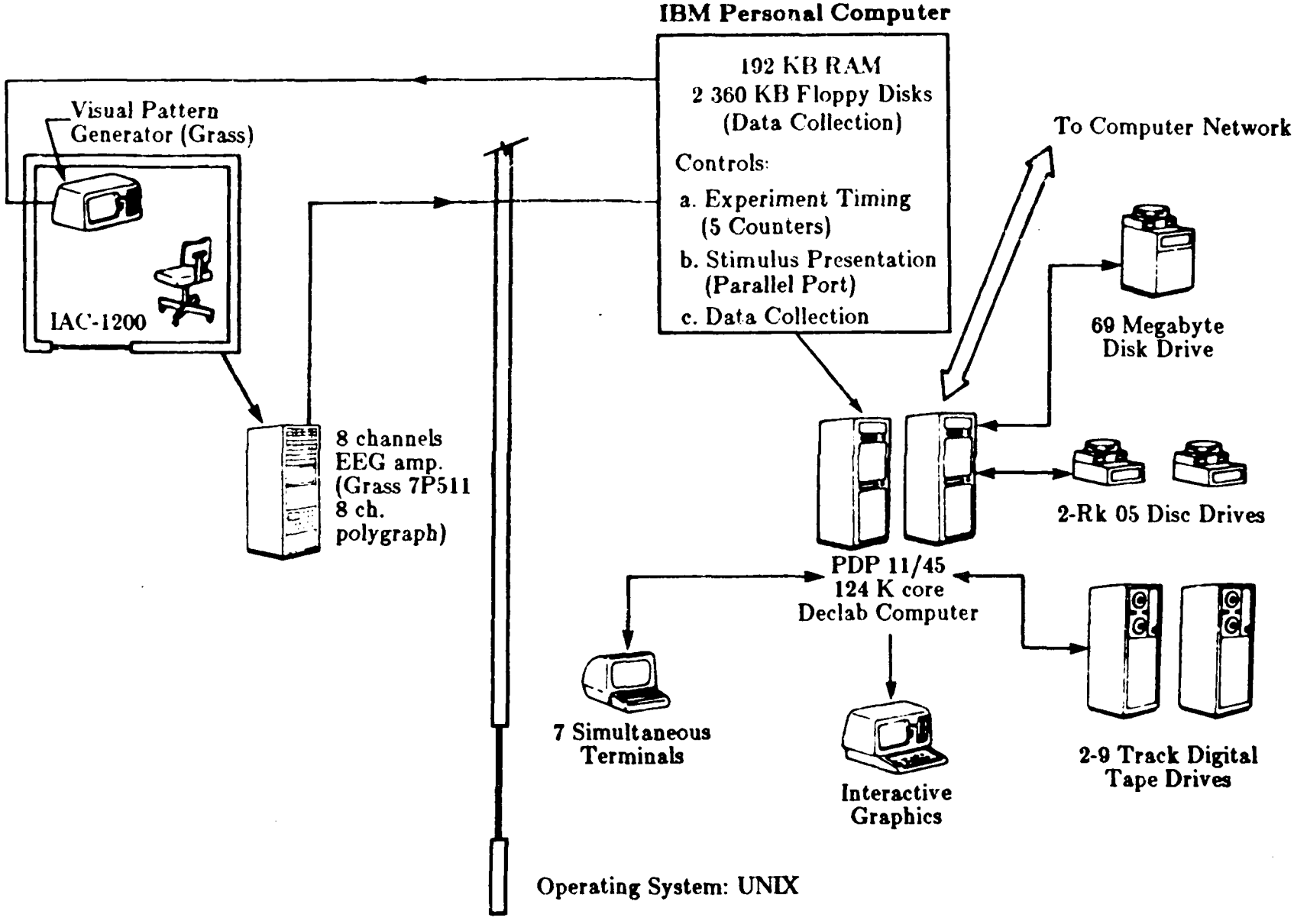


Figure 6-1 EEG Signal Processing Laboratory Facilities for Data Acquisition and Analysis

can be triggered automatically under the control of one of the 16 bit counters or manually under the control of software. Data are stored temporarily in RAM memory during sampling and transferred to a floppy disk file during the interstimulus interval.

6.2. Experiment Design and Control

A versatile set of assembly language programs has been written to perform specific functions at maximum efficiency. These subroutines were designed to be called from BASIC programs:

(1) initialization of the TECMAR board,

(2) calibration of the Grass EEG amplifiers,

(3) A/D conversion of amplified and filtered multielectrode EEG data,

(4) rapid transfers of sampled data to disk files during an experiment, and

(5) on-line plots of sampled multielectrode EEG data on the PC video monitor.

Simple BASIC programs can be written to design an EP experiment, obtain timing and stimulus parameters from the user, and then call specific functions (in assembly language) to run the experiment. The combination of BASIC user interface and assembly language drivers provides a powerful system for rapid development of versatile EP experiments. Facilities are also available for programming in C and FORTRAN.

6.3. Capabilities

Besides being the basis of a powerful data acquisition system, the PC can also perform on-line filtering, detection, and classification of sampled data using the capabilities of the 8087 hardware math coprocessor. By significantly increasing the speed of the math operations, the on-line processing can be performed in virtually real time.

The PC can also simultaneously control several instruments now present in the EEG Signal Processing Lab using the individual bits in the parallel

ports:

(1) Grass Pattern Generator for checkerboard patterned stimulation,

(2) a video monitor for more complex patterns (figures, numbers, or letters),

(3) auditory equipment for auditory evoked potentials,

(4) a set of LED goggles for flash evoked potentials, and

(5) an on-line averager for monitoring EPs over the course of an experiment.

The PC-based data acquisition system offers the advantages of versatility (easy to design and modify experiments), ease of operation (any user with little training can run the experiments), and low cost. Future developments include the purchase of an IBM XT with a 100 Mbyte hard disk or the faster and more powerful IBM AT. On-line signal processing software will also be developed to perform near real time detection, estimation, and classification of EPs.

7. PROJECT PERSONNEL

Clare D. McGillem, Ph.D., Professor of Electrical Engineering

Jorge I. Aunon, Sc.D., Professor of Electrical Engineering

8. DEGREES GRANTED

Jeffrey M. Moser, Ph.D. Dissertation: "Classification of Single Evoked Brain Potentials Using Time-Frequency Amplitude Features," August 1984.

John J. Westerkamp, Ph.D. Dissertation: "A Multichannel Time-Varying Filter for Estimating Nonstationary Signals in Noise," August 1985.

9. PUBLICATIONS

9.1. Journal Articles and Conference Presentations

J. I. Aunon and C. D. McGillem, "Event Related Potential Estimation Procedures," AIMI Conf., Houston, TX, May 1983.

J. J. Westerkamp, J. I. Aunon, and C. D. McGillem, "Classification of Event Related Potentials," *Proceedings of the 1983 IEEE* Frontiers of Engineering and Computing in Health Care Conference," Columbus, OH, Sept. 1983.

K. B. Yu and C. D. McGillem, "Optimum Filters for Estimating Evoked Potential Waveforms," *IEEE Transactions on Biomedical Engineering,* BME-30, No. 11, Nov. 1983.

J. J. Westerkamp, J. I. Aunon, C. D. McGillem, and J. M. Moser, "Sequential Detection of Changes in Evoked Potentials," *Proceedings of the 1984 IEEE Frontiers of Engineering and* Computing in Health Care Conference," Los Angeles, CA, Sept. 1984.

J. I. Aunon, J. J. Westerkamp, C. D. McGillem, and K. Fortier, "Prony Decomposition of Evoked Potentials," *Proceeding of the 1984 IEEE Frontiers of Engineering and* Computing in Health Care Conference," Los Angeles, CA, Sept. 1984.

C. D. McGillem, J. I. Aunon, and C. A. Pomalaza, "Improved Waveform Estimation Procedures for Event Related Potentials," *IEEE Transactions on Biomedical Engineering,* BME-32, No. 6, June 1985.

D. L. Halliday, C. D. McGillem, J. Westerkamp, and J. I. Aunon, "Optimal and Suboptimal Feature Selection for Classification of Evoked Brain Potentials," *IEEE Transactions on Systems, Man, and Cybernetics,* SMC-15, No. 3, May 1985.

9.2. Pending Publications

C. D. McGillem, J. I. Aunon, and K. B. Yu, "Signals and Noise in Evoked Brain Potentials," submitted to *IEEE Transactions on Biomedical Engineering.*

C. D. McGillem and J. I. Aunon, "Analysis of Event Related Potentials," Chapter 5 in *Handbook of Electroencephalography and Clinical Neurophysiology,* A. Gevins and A. Remond, Eds., Elsevier, Amsterdam.

J. J. Westerkamp, J. I. Aunon, and C. D. McGillem, "A Multichannel Time-Varying Filter for Estimating Evoked Potentials," submitted to the *1985 IEEE Frontiers of Engineering and Computing in Health Care* Conference," Chicago, IL, Sept. 30 - Oct. 2, 1985.

9.3. Proposed Publications

J. J. Westerkamp and J. I. Aunon, "An Implementation Method for *A Posteriori* Estimation Filters," will be submitted to *IEEE Transactions on Acoustics Speech and Signal Processing.*

J. J. Westerkamp and J. I. Aunon, "Optimum Multichannel *A Posteriori* Estimates of Single Response Evoked Potentials," will be submitted to *IEEE Transactions on Biomedical Engineering.*

LIST OF REFERENCES

LIST OF REFERENCES

Allen1977.
Allen, J. B., "Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 25, no. 3, 1977.

Aunon1977.
Aunon, J. I. and McGillem, C. D., "High Frequency Components in the Spectrum of the Visual Evoked Potential," *Journal of Bioengineering*, vol. 1, pp. 157-164, 1977.

Aunon1979.
Aunon, J. I. and McGillem, C. D., "Detection and Processing of Individual Components in the VEP," *Psychophysiology*, vol. 16, no. 1, pp. 71-79, 1979.

Aunon1982a.
Aunon, J. I., McGillem, C. D., and O'Donnell, R. D., "Comparison of Linear and Quadratic Classification of Event-Related Potentials on the Basis of Their Exogenous or Endogenous Components," *Psychophysiology*, vol. 19, no. 5, pp. 531-537, 1982.

Aunon1982b.
Aunon, J. I. and McGillem, C. D., "On the Classification of Single Evoked Potentials Using a Quadratic Classifier," *Computer Programs in Biomedicine*, vol. 14, 1982.

Blackman1958.
Blackman, R. B. and Tukey, J. W., *The Measurement of Power Spectra*, Dover, New York, 1958.

Booton1952.
Booton, R. C., "An Optimization Theory for Time-Varying Linear Systems with Nonstationary Statistical Inputs," *Proc. IRE*, vol. 40, pp. 977-981, 1952.

Childers1982.
Childers, D. G., Bloom, P. A., Arroyo, A. A., Roucos, S. E., Fischler, I. S., Achariyapaopan, T. and Perry Jr., N. W., "Classification of Cortical Responses Using Features from Single EEG Records," *IEEE Trans. Biomedical Eng.*, vol. 29, no. 6, 1982.

Cooper1965.
Cooper, R., Winter, A. L., Crow, H. J., and Walter, W. G., "Comparison of the Subcortical, Cortical, and Scalp Activity Using Chronically Indwelling Electrodes," *Electroen. Clin. Neuroph.*, vol. 18, pp. 217-228, 1965.

Creutzfeldt1966.
Creutzfeldt, O. D., Watanabe, S. and Lux, H. D., "Relations Between EEG Phenomena and Potentials of Single Cortical Cells (Parts I and II)," *Electroenceph. Clin. Neurophys.*, vol. 20, 1966.

Dixon1975.
Dixon, W. J., *BMDP, Biomedical Computer Programs*, UCLA Health Science Computing Facility, Los Angeles, 1975.

Donchin1966.
Donchin, E., "A Multivariate Approach to the Analysis of Averaged Evoked Potentials," *IEEE Trans. Biomedical Eng.*, vol. 13, pp. 131-139, 1966.

Donchin1975.
Donchin, E. and Herning, R. I., "A Simulation Study of the Efficacy of Stepwise Discriminant Analysis in the Detection and Comparison of Event Related Potentials," *Electroen. Clin. Neuroph.*, vol. 38, pp. 51-68, 1975.

Franks1981.
Franks, L. E., *Signal Theory*, Dowden & Culver, Stroudsburg, PA, 1981.

Fukunaga1970.
Fukunaga, K. and Koontz, W. L., "Application of Karhunen-Loeve Expansion to Feature Selection and Ordering," *IEEE Trans. Computers*, vol. 19, 1970.

Fukunaga1971.
Fukunaga, K. and Kessel, D. L., "Estimation of Classification Error," *IEEE Trans. Computers*, vol. 20, 1971.

Fukunaga1972.
Fukunaga, K., *Introduction to Statistical Pattern Recognition*, Academic Press, New York, 1972.

Halliday1985.
Halliday, D. L., McGillem, C. D., Westerkamp, J., and Aunon, J. I., "Optimal and Suboptimal Feature Selection for Classification of Evoked Brain Potentials," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 15, no. 3, 1985.

Hanson1971.
Hanson, R. J., "A Numerical Method for Solving Fredholm Integral Equations of the First Kind Using Singular Values," *SIAM J. Numer. Anal.*, vol. 8, no. 3, pp. 616-622, 1971.

Hughes1968.
Hughes, G. F., "On the Mean Accuracy of Statistical Pattern Recognizers," *IEEE Trans. Info. Theory*, vol. 14, no. 1, 1968.

Jansen1981a.
Jansen, B. H., Bourne, J. R. and Lenten, R., "Piece-wise Analysis: An Objective Evaluation," *Int. J. Biomedical Computing*, vol. 12, 1981.

Jansen1981b.
Jansen, B. H., Bourne, J. R. and Ward, J. W., "Autoregressive Estimation of Short Segment Spectra for Computerized EEG Analysis," *Int. J. Biomedical Computing*, vol. 12, 1981.

Jasper1958.
Jasper, Herbert H., "The Ten Twenty Electrode System of the International Federation," *Electroen. Clin. Neuroph.*, vol. 10, pp. 371-375, 1958.

Jervis1983.
Jervis, B. W., Nichols, M. J., Johnson, T. E., Allen, E. and Hudson, N. R., "A Fundamental Investigation of the Composition of Auditory Evoked Potentials," *IEEE Trans. Biomedical Engineering*, vol. 30, no. 1, 1983.

John1973.
John, E. R., Walker, P., Cawood, D., Rush, M. and Gehrmann, J., "Factor Analysis of Evoked Potentials," *Electroen. Clin. Neuroph.*, vol. 34, 1973.

Kanal1974.
Kanal, L., "Patterns in Pattern Recognition: 1968-1974," *IEEE Trans. Info. Theory*, vol. 20, no. 6, 1974.

Kazakos1982a.
Kazakos, D., "Signal Detection Under Mismatch," *IEEE Trans. Info. Theory*, vol. 28, no. 4, 1982.

Kazakos1982b.
Kazakos, D., "Statistical Discrimination Using Inaccurate Models," *IEEE Trans. Info. Theory*, vol. 28, no. 5, 1982.

Levinson1946.
Levinson, N., "The Wiener RMS Error Criterion in Filter Design and Prediction," *Journal of Mathematics and Physics*, vol. 25, pp. 261-278, 1946.

Lissack1976.
Lissack, T. and Fu, K. S., "Parametric Feature Extraction Through Error Minimization Applied to Medical Diagnosis," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 6, no. 9, 1976.

McGillem1977.
McGillem, C. D. and Aunon, J. I., "Measurements of Signal Components in Single Visually Evoked Potentials," *IEEE Trans. Biomedical Engineering*, vol. 24, no. 3, 1977.

McGillem1981.
McGillem, C. D., Aunon, J. I. and Childers, D. G., "Signal Processing in Evoked Potential Research: Applications of Filtering and Pattern Recognition," in *Critical Reviews in Bioengineering*, pp. 225-265, CRC, Cleveland, 1981.

McGillem1983.
Aunon, J. I., "New Techniques for Measuring Single Event Related Brain Potentials," *Technical Report AFOSR-TR-82-0901*, 1983.

McGillem1985.
McGillem, C. D., Aunon, J. I. and Pomalaza, C. A., "Improved Waveform Estimation for Event Related Potentials," *IEEE Trans. Biomedical Engineering*, vol. 32, pp. 371-379, 1985.

Morf1977.
Morf, M., Dickinson, B., Kailath, T., and Vieira, A., "Efficient Solution of Covariance Equations for Linear Prediction," *IEEE Trans. Acoust. Speech and Signal Proc.*, vol. 25, no. 5, pp. 429-433, 1977.

Moser1980.
Moser, J. M., "Decomposition of the Electroencephalogram," in *Master Thesis*, Purdue University, West Lafayette, IN, 1980.

Moser1982.
Moser, J. M., Aunon, J. I., and McGillem, C. D., "Classification of Event-Related Brain Potentials," *Proc. 35th ACEMB Conf.*, 1982.

Moser1984.
Moser, J. M., "Classification and Detection of Single Evoked Brain Potentials Using Time-Frequency Amplitude Features," in *PhD Dissertation*, Purdue University, West Lafayette, 1984.

Moukas1982.
"Automatic Identification of Noise Pollution Sources," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 12, no. 5, 1982.

Mountcastle1974.
Mountcastle, V. B., *Medical Physiology, Thirteenth Edition*, C. V. Mosky, St. Louis, 1974.

Mucciardi1971.
Mucciardi, A. N. and Gose, E. E., "A Comparison of Seven Techniques for Choosing Subsets of Pattern Recognition Properties," *IEEE Trans. Computers*, vol. 20, 1971.

Pomalaza1979.
Pomalaza, C., Aunon, J. I., McGillem, C. D., "Nonstationary EEG Analysis Using the Maximum Entropy Method," *Proceedings 1979 IEEE Frontiers of Engineering and Computing in Health Care Conference*, 1979.

Portnoff1976.
Portnoff M. R., "Implementation of the Digital Phase Vocoder Using the Fast Fourier Transform," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 23, 1976.

Sanderson1980.
Sanderson, A. C., Segen, J. and Richey E., "Hierarchical Modeling of EEG Signals," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 2, no. 5, 1980.

Schafer1973.
Schafer, R. W. and Rabiner, L. R., "Design and Simulation of a Speech Analysis-Synthesis System Based on Short-Time Fourier Analysis," *IEEE Trans. Audio Electroacoust.*, vol. 21, 1973.

Sencaj1979.
Sencaj, R. W., Aunon, J. I., and McGillem, C. D., "Discrimination Among Visual Stimuli By Classification of Their Single Evoked Potentials," *Med. Biol. Engr. Comp.*, vol. 17, pp. 391-396, 1979.

Slakter1965.
Slakter, M. J., "A Comparison of the Pearson Chi-Square and Kolmogorov Goodness-of-Fit Tests With Respect to Validity," *Journal of American Statistical Association*, vol. 60, pp. 854-858, 1965.

Sternberg1966.
Sternberg, S., "High-Speed Scanning in Human Memory," *Science*, vol. 153, 1966.

Strang1976.
Strang, G., *Linear Algebra and Its Applications*, Academic Press, New York, 1976.

Sullivan1984.
Sullivan, B. J. and Liu, B., "On the Use of Singular Value Decomposition and Decimation in Discrete-Time Band-Limited Signal Extrapolation," *IEEE Trans. Acoust. Speech and Signal Proc.*, vol. 32, no. 6, p. 1201-1212, 1984.

Tanaka1979.
"A Dynamic Processing Approach to Phoneme Recognition (Part I) -- Feature Extraction," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 27, no. 6, 1979.

VanHoek1974.
Van Hoek, L. D., "Multivariate Vectorial Analysis of the Visual Evoked Response," *Kybernetik*, vol. 15, 1974.

VanTrees1968.
VanTrees, H. L., *Detection, Estimation, and Modulation Theory*, Wiley, New York, 1968.

Varah1973.
Varah, J. M., "On the Numerical Solution of Ill-Conditioned Linear Sytems with Applications to Ill-Posed Problems," *SIAM J. Numer. Anal.*, vol. 10, no. 2, pp. 257-267, 1973.

Vidal1977.
Vidal, J. J., "Real Time Detection of Brain Events in EEG," *Proc. IEEE*, vol. 65, no. 5, pp. 633-641, 1977.

Westerkamp1985.
Westerkamp, J. J., "A Multichannel Time-Varying Filter for Estimating Nonstationary Signals in Noise," in *PhD Dissertation*, Purdue University, West Lafayette, 1985.

White1976.
White, G. M. and Neely, R. B., "Speech Recognition Experiments With Linear Prediction, Bandpass Filtering, and Dynamic Programming," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 24, no. 2, 1976.

Wilkinson1963.
Wilkinson, J. H., *Rounding Errors in Algebraic Processes*, Prentice-Hall, Englewood Cliffs, 1963.

Williams1950.
Williams, C. A., "On the Choice of the Number and Width of Classes for Chi-Square Test of Goodness-of-Fit," *Journal of the American Statistical Association*, vol. 45, pp. 77-86, 1950.

Wolf1959.
Wolf, J. K., "On the Detection and Estimation Problem for Multiple Nonstationary Random Processes," in *PhD Dissertation*, Princeton University, Princeton, NJ, 1959.

Yu1982.
Yu, K., "Optimum Time-Varying Filters for Transient Waveform Analysis," in *PhD Dissertation*, Purdue University, W. Lafayette, 1982.

Yu1983.
Yu, K. and McGillem, C. D., "Optimum Filters for Estimating Evoked Potential Waveforms," *IEEE Trans. Biomedical Eng.*, vol. 30, no. 11, pp. 730-737, 1983.

APPENDICES

Appendix A: Classifier Evaluation Using Simulated Data

The artificial data sets were generated by summing simulated signals and ongoing EEG. The signals were generated by several methods based on different ways of modeling data. Data sets were generated which contained different signals added to the ongoing EEG (noise) at various signal-to-noise ratios (SNRs). These data sets were processed by the algorithms described in Section 2.

The signal represents the brain generated waveform in response to the applied stimulus. The first signal used was an averaged EP taken from human data. The EP was windowed with a Tukey Window and added to the noise. The second signal was composed of a summation of basic components of the same form with arbitrarily chosen amplitudes and latencies. The signal was the same in each generated record. The representation of this signal $s_2(t)$ is

$$s_2(t) = \sum_{i=1}^{M} A_i \cdot f(t - T_i), \qquad \text{(A-1)}$$

where A_i is the amplitude of the i^{th} component,

$f(t)$ is the functional form of the component,

T_i is the time location of the i^{th} component,

and M is total number of components in the signal.

Two forms of the component $f(t)$ were tested. The first was a Gaussian pulse of the form

$$f_1(t) = \exp\left[\frac{(t - T_i)^2}{2\sigma^2}\right], \qquad \text{(A-2)}$$

where σ represents the width. The value of σ was set to 15 ms, a typical value found in past research (McGillem 1977). The second component was a sinusoidal pulse or a windowed sinusoid. The window function chosen was a raised cosine window, the same as that chosen in forming the filter in the 2 step process. The equation of this component $f_2(t)$ is:

$$f_2(t) = \cos[2\pi f(t-T_i)]\cdot\left\{\frac{1}{2}+\frac{1}{2}\cos\left[\frac{\pi}{a}(t-T_i)\right]\right\}, \text{for } T_i-a\leq t\leq T_i+a \qquad \text{(A-3)}$$

$$= 0, \text{ otherwise}$$

where f = the sinusoidal frequency,

and a = one-sided width of window in seconds.

The background noise used for this data was human ongoing EEG data which also included instrumentation noise. The known signal was added to this data with various signal to ongoing EEG power ratios. The background noise data records were first normalized by removing the estimated mean of each data record and then scaling to make the root-mean-square (RMS) amplitude =1. The signal was also normalized by scaling to make the RMS amplitude =1. The mean of the signal was not removed to preserve polarity asymmetries which may be characteristic of the EP signal. The amplitude of the signal was then adjusted before being added to noise to produce data records with predetermined SNR's.

Four types of tests signals were used. The first used the averaged EP's from subject 2. The portion of the EP from 0 to 500 ms was used with the end 100 ms of data windowed by a raised cosine to provide smooth transitions at its ends when it was added to the noise. The t1 (target set, 1 target in set) and n1 (nontarget set 1 target in set) sets were used. The 1 second segments of the data records from subject 2 before stimulation were normalized and the signal added to give SNR's of 0dB, -6dB, and -12dB for all of the generated data sets. These values were chosen because they bound the estimated SNR of the actual EP data. Figure A-1 portrays these two data sets. Four generated data records and the signal are displayed for the t1 data (top) and n1 data (bottom). The SNR in these data records as in all of the following artificially generated data records is -6dB.

Figure A-2 portrays EP's composed of 15 Gaussian pulse components embedded in human EEG data. Eight of the components in the signals in data set 2 are the same as in data set 1. The other 7 components have the same amplitudes but increased latencies with respect to those components in the signals of data set 1. A component's latency shift was made proportional to its latencies so that latter components were shifted further. The largest latency component shifted 72 ms. The amplitudes of the corresponding components in

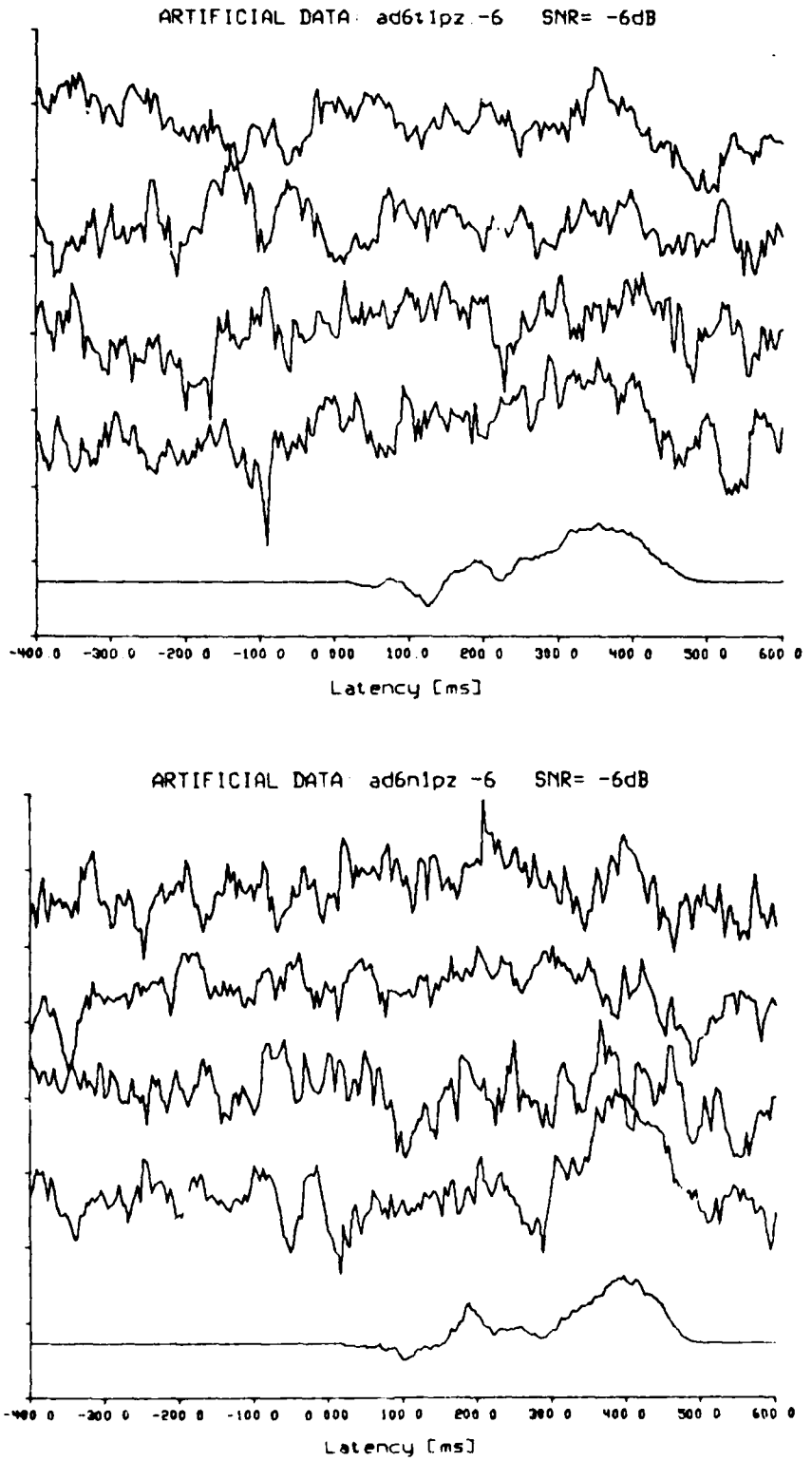


Figure A-1 Examples of Artificial Data Records and the Signals for Averaged Evoked Potentials Embedded in On-Going EEG, SNR = -6 dB

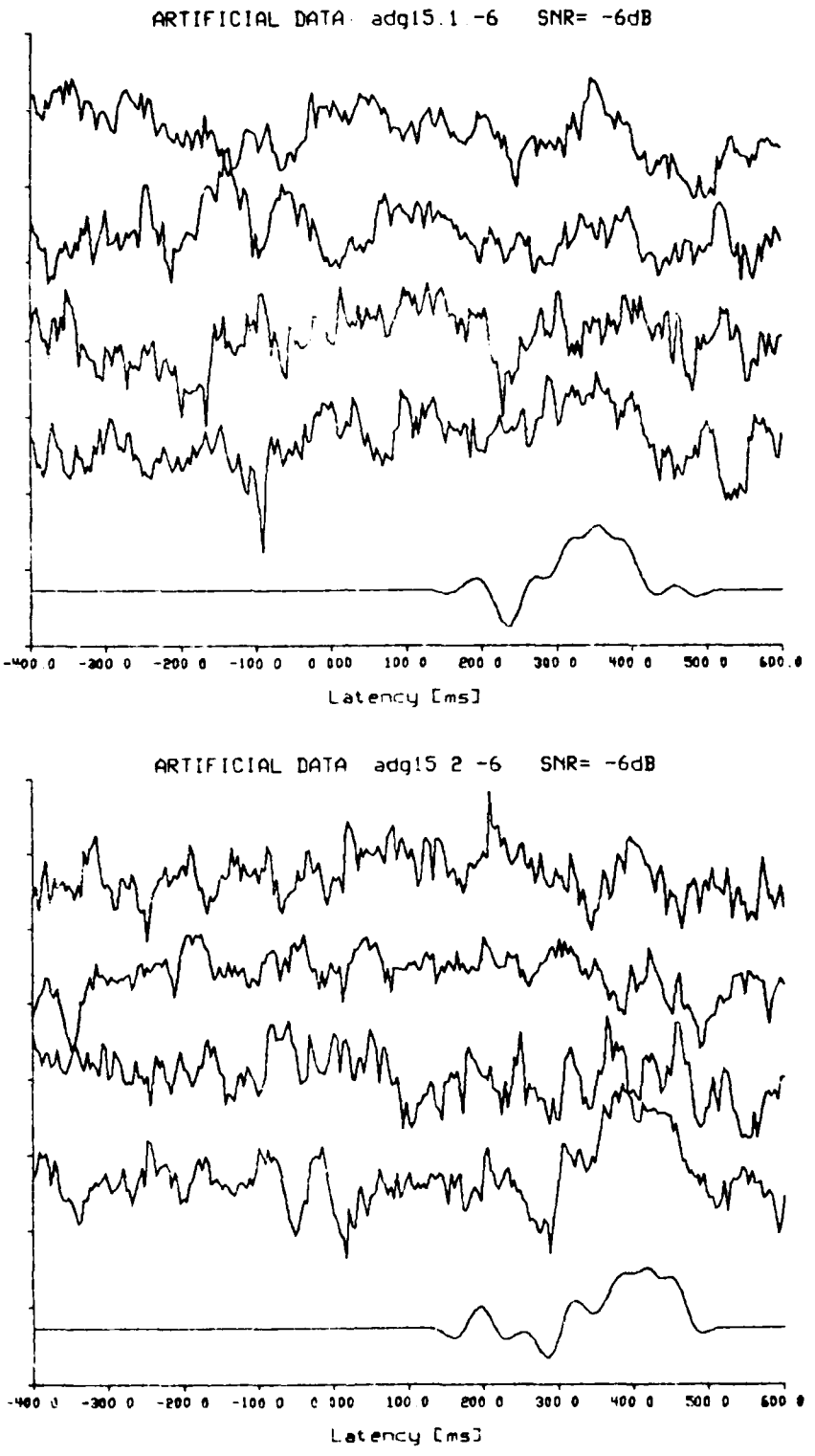


Figure A-2 Examples of Artificial Data Records and the Signals for Signals Composed of Gaussian Components Embedded in On-Going EEG, SNR = -6 dB

these data sets remained the same, only the latencies of the components were increased.

Another type of signal was also generated which was composed of windowed sinusoids. Two sets of data with different signals composed of just one sinusoid were generated. Both data sets used a 400 ms wide raised cosine to window different arbitrarily selected frequency sinusoids with cosine phase of zero. The latencies in each data set were identical. Figure A-3 portrays these signals and the signals embedded in EEG data.

Two classes of signals composed of 3 windowed sinusoids of different frequencies and latencies were generated. The second class had the same type of components with the same frequencies and amplitudes but 2 components had increased latencies, They were embedded in EEG data and are portrayed in Figure A-4.

The artificial data sets were processed in the same manner as the human data. Tables A-1 through A-4 list the accuracies and number of selected features for the data containing the 4 types of signals generated. In all of these tables, each column lists the results for a different SNR for processing with or without frequency transformation. The upper third contains the results for detection of the first data set and the middle third the results for detection of the second set. The lower third contains the results for classification between the two sets. Table A-5 summarizes the results of processing the artificial data. The table entries are the highest accuracies achieved for processing either with or without frequency transformation. The entries in all of these tables for raw features used in detection are averages of four runs. Each run used different segments of ongoing EEG prior to stimulation for the class 2 data. This resulted in slightly different results for each run due to the different data.

Data sets composed of windowed averaged EP's in EEG were tested (Table A-1). Detection at SNR's of 0dB produced very high accuracies for all types of features, 99% and 99.1% for 6-8 raw features and 99.5% for 5 frequency features for the two data sets. Filtered features yielded 100% accuracy with 3 features. For classification between the two signals, 9 raw features yielded 85.7% accuracy and frequency features yielded up to 91.8% at 5 features. Filtered features increased these accuracies to 95.9% for 8 features, a 10.2% improvement.

Detection at SNR's of -6dB yielded accuracies of 89.9% and 92.5% for 7 raw features, and up to 89.3% for 6 frequency features and 93.5% for 8 features for the two data sets. Filtered features produced increased accuracies of up to 99% and 100% for 8 and 5 features. For classification between the

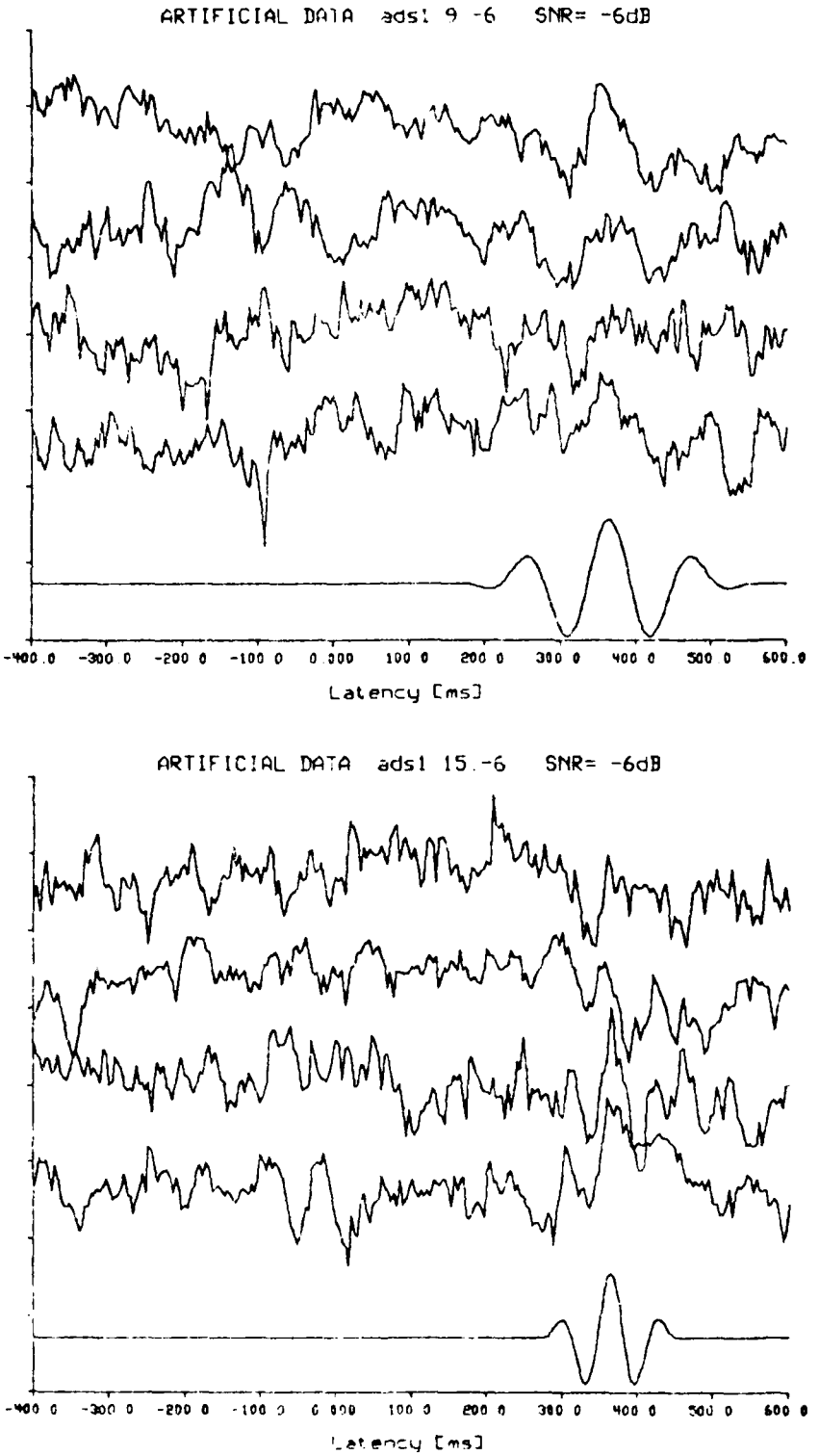


Figure A-3 Examples of Artificial Data Records and the Signals for Signals Composed of a Single Windowed Sinusoid Embedded in On-Going EEG, SNR = -6 dB

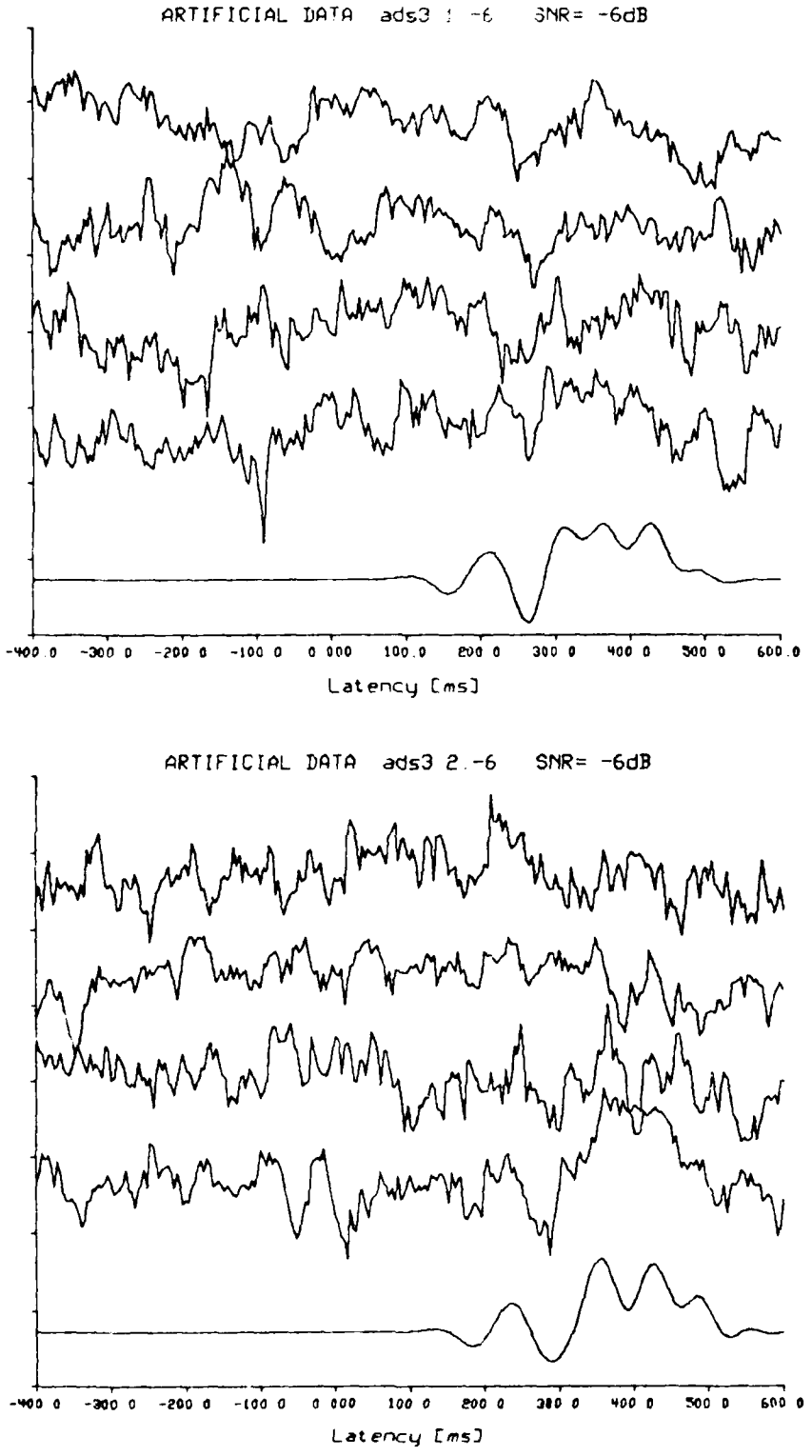


Figure A-4 Examples of Artificial Data Records and the Signals for Signals Composed of Three Windowed Sinusoids Embedded in On-Going EEG, SNR = -6 dB

Table A-1 Results From the 2-Step Classifier/Feature Selection Process for Artificial Data Composed of Windowed Averaged Evoked Potentials in EEG. Table Entries: Percent Classification Accuracy / Number of Features

Frequency transformation→		Without			With		
Data	BW↓ SNR→	O dB	-6 dB	-12 dB	O dB	-6 dB	-12 dB
Detect 1	Raw	99/7.5	89.8/7	78.8/5	99/7.5	89.8/7	78.8/5
	Frequency	99.5/5	89.3/6	82.1/6	98.5/6	86.7/4	78.1/6
	5 Hz	99/4	92.4/5	96.4/9	98/6	90.8/10	78.6/8
	10 Hz	100/4	92.9/3	97.5/4	97.5/2	91.3/7	88.8/10
	15 Hz	100/3	95.9/8	98.5/5	100/5	93.9/9	96.9/10
	X0.25 Fo	99/3	92.4/10	98.5/8	98.5/5	89.8/6	80.1/4
	X0.5 Fo	99.5/4	91.8/5	83.7/5	99/8	91.3/10	98.5/8
	X0.75 Fo	98.5/3	91.8/9	85.7/6	99/4	92.9/9	96.9/10
	X1.0 Fo	99.5/4	95.4/7	88.3/8	97.5/4	99/8	97.5/10
	Best	15 Hz	15 Hz	15 Hz	15 Hz	X1.0	X0.5
Detect 2	Raw	99.1/6	92.5/7	83.5/6	99.1/6	92.5/7	83.5/6
	Frequency	99.5/5	90.5/9	85.5/7	98.5/4	93.5/8	83/7
	5 Hz	100/3	96.5/5	89/8	98.5/3	92.5/4	84/5
	10 Hz	99.5/2	98/6	93.5/9	98.5/3	99.5/8	83/9
	15 Hz	100/3	100/5	98/10	99/3	98.5/7	87/8
	X0.25 Fo	99/3	96/8	98/10	97.5/3	93.5/6	87.5/6
	X0.5 Fo	99.5/3	94.5/5	97.5/8	99/4	91/3	90/10
	X0.75 Fo	100/4	98/7	99.5/8	98/3	93.5/10	98/10
	X1.0 Fo	99.5/3	99/8	98.5/8	97.5/2	97/7	97.5/10
	Best	Several	15 Hz	X.75	15 Hz	10 Hz	X0.75
Classify 1 vs. 2	Raw	85.7/9	80.1/8	71.4/5	85.7/9	80.1/8	71.4/5
	Frequency	91.8/5	83.2/5	73/3	92.4/10	83.2/8	72.5/4
	5 Hz	90.3/6	85.2/4	75/3	91.8/8	85.7/5	75.5/2
	10 Hz	89.8/4	81.6/5	74/5	92.9/5	85.7/5	79.1/8
	15 Hz	88.3/7	81.1/5	74/7	93.9/10	85.2/9	71.9/8
	X0.25 Fo	92.4/5	86.2/8	75.5/5	91.8/6	86.2/4	77/4
	X0.5 Fo	91.8/6	85.7/5	77/9	95.9/8	83.7/8	77/4
	X0.75 Fo	93.4/5	83.2/4	80.1/8	93.4/9	87.8/7	79.6/10
	X1.0 Fo	94.4/8	83.7/10	79.6/9	90.3/9	83.2/6	76.5/7
	Best	X1.0	X.25	X.75	X0.5	X0.75	X0.75

Table A-2 Results From the 2-Step Classifier/Feature Selection Process for Artificial Data Composed of 15 Gaussian Components in EEG. Table Entries: Percent Classification Accuracy / Number of Features.

Frequency transformation→		Without			With		
Data	BW↓ SNR→	O dB	-6 dB	-12 dB	O dB	-6 dB	-12 dB
Detect 1	Raw	98.5/6.5	83.6/5	66.2/6	98.5/6.5	83.6/5	66.2/6
	Frequency	99/6	85.7/8	73/5	98/4	85.7/6	67.9/2
	5 Hz	98/2	83.2/5	98/10	99/4	85.7/7	76.5/8
	10 Hz	98.5/2	98/10	98/9	100/4	91.8/10	66.3/3
	15 Hz	99/3	99.5/8	95.4/8	100/4	100/9	86.7/10
	X0.25 Fo	99/3	80.6/2	95.4/10	97.5/2	86.2/4	76/10
	X0.5 Fo	99.5/4	86.7/10	98/7	97.5/3	90.8/9	70.9/10
	X0.75 Fo	98.5/2	88.8/10	95.9/7	100/4	87.8/8	65.8/6
	X1.0 Fo	98.5/2	96.9/9	98/9	99.5/6	85.7/5	65.3/2
	Best	X.50	15 Hz	X.50	Several	15 Hz	15 Hz
Detect 2	Raw	96.1/8	78.5/6	70.6/6	96.1/8	78.5/6	70.6/6
	Frequency	100/6	80/7	71.5/6	99/8	81/8	74/3
	5 Hz	99.5/4	92.5/9	94/10	99.5/4	86.5/4	72/9
	10 Hz	98.5/3	95.5/9	80/10	98/2	88.5/9	97/10
	15 Hz	99.5/5	94/10	96.5/9	98.5/3	96/10	98/8
	X0.25 Fo	99/2	87.5/9	70/5	100/4	85.5/3	93.5/10
	X0.5 Fo	99.5/2	89/10	83/10	99/5	85.5/10	95.5/10
	X0.75 Fo	99.5/3	98/10	92/10	98/2	88/6	97/10
	X1.0 Fo	99.5/4	95/8	92/8	100/4	91.5/6	96.5/9
	Best	Freq.	X.75	15 Hz	Several	15 Hz	15 Hz
Classify 1 vs. 2	Raw	89.8/5	77.6/6	63.8/5	89.8/5	77.6/6	63.8/5
	Frequency	100/4	86.2/6	74/8	100/4	84.7/4	73/8
	5 Hz	100/2	90.3/8	74.5/6	99.5/2	87.2/9	75.5/5
	10 Hz	99.5/2	87.8/5	76.5/8	99.5/2	87.2/6	75/9
	15 Hz	99.5/6	85.7/7	69.4/5	99/6	84.7/5	63.3/8
	X0.25 Fo	100/2	86.7/4	72.5/3	99.5/3	89.3/9	77/6
	X0.5 Fo	100/2	87.2/7	75.5/6	99.5/3	87.2/5	72/5
	X0.75 Fo	100/2	85.7/4	74.5/5	99.5/3	86.2/9	70.4/7
	X1.0 Fo	100/2	89.3/9	73/5	99.5/2	86.2/10	71.9/9
	Best	Several	5 Hz	10 Hz	Freq.	X0.25	X0.25

Table A-3 Results From the 2-Step Classifier/Feature Selection Process for Artificial Data Composed of 1 Windowed Sinusoid in EEG. Table Entries: Percent Classification Accuracy / Number of Features.

Frequency transformation→		Without			With		
Data	BW↓ SNR→	O dB	-6 dB	-12 dB	O dB	-6 dB	-12 dB
Detect 1	Raw	98.1/5	84.7/7.5	73.4/4.5	98.1/4	84.7/7.5	73.4/4.5
	Frequency	100/2	92.9/4	75/4	100/3	91.8/4	79.1/5
	5 Hz	100/2	97.5/5	81.1/4	100/2	95.4/3	99.5/10
	10 Hz	100/2	99.5/8	96.4/8	100/1	96.9/7	99.5/10
	15 Hz	100/2	98.5/7	98/7	99/1	98.5/7	99.5/6
	X0.25 Fo	99.5/2	100/7	91.3/8	100/2	95.4/6	99/9
	X0.5 Fo	100/2	100/7	90.8/7	100/2	94.4/6	96.4/7
	X0.75 Fo	99.5/2	95.4/5	93.4/8	100/1	95.9/7	99.5/10
	X1.0 Fo	100/2	96.4/8	98/8	100/1	93.9/4	98.5/7
	Best	Several	Several	15 Hz	Several	15 Hz	15 Hz
Detect 2	Raw	99/4.5	88.5/7	74.4/5	99/4.5	88.5/7	74.4/5
	Frequency	99.5/3	91/6	71.5/5	100/5	90.5/6	72.5/7
	5 Hz	99.5/4	98/8	84/8	99/4	98.5/9	77.5/10
	10 Hz	99.5/4	99/8	88.5/8	100/3	100/8	93/9
	15 Hz	99.5/2	100/5	93/8	100/2	99.5/7	99.5/8
	X0.25 Fo	99.5/5	93.5/6	85.5/7	97.5/2	98.5/10	75.5/3
	X0.5 Fo	97.5/3	99/7	98/8	99.5/2	97/7	82.5/9
	X0.75 Fo	98.5/3	99.5/6	98.5/7	100/3	100/7	100/9
	X1.0 Fo	99/2	99/5	98/8	100/2	98/5	99/8
	Best	15 Hz	15 Hz	X.75	Several	X0.75	X0.75
Classify 1 vs. 2	Raw	99.5/4	83.7/10	71.9/3	99.5/4	83.7/10	71.9/3
	Frequency	100/3	95.9/8	76.5/7	100/3	93.9/7	78.1/6
	5 Hz	99.5/3	95.9/6	76.5/5	99.5/2	94.4/5	76.5/3
	10 Hz	100/2	95.4/7	82.7/8	100/2	95.9/5	78.6/3
	15 Hz	100/3	95.9/4	81.1/3	100/2	95.9/7	78.1/5
	X0.25 Fo	99.5/4	92.4/3	75.5/8	100/3	93.4/5	73.5/2
	X0.5 Fo	99.5/2	94.9/6	76.5/3	100/2	95.9/6	76/4
	X0.75 Fo	100/2	94.9/5	80.6/4	100/2	97.5/7	78.6/5
	X1.0 Fo	100/2	95.9/4	81.1/3	100/3	96.4/6	79.1/4
	Best	Several	Several	10 Hz	Several	X0.75	X1.0

Table A-4 Results From the 2-Step Classifier/Feature Selection Process for Artificial Data Composed of 3 Windowed Sinusoids in EEG. Table Entries: Percent Classification Accuracy / Number of Features.

Frequency transformation→		Without			With		
Data	BW↓ SNR→	O dB	-6 dB	-12 dB	O dB	-6 dB	-12 dB
Detect 1	Raw	99.4/7	86.2/9	73.9/5	99.4/7	86.2/9	73.9/5
	Frequency	100/7	84.7/5	68.4/5	99.5/4	87.8/6	74/7
	5 Hz	100/4	88.3/8	68.9/7	99/3	87.8/6	69.9/5
	10 Hz	100/2	92.4/8	76.5/6	100/2	96.4/10	81.6/5
	15 Hz	100/4	95.4/10	81.1/10	100/3	93.4/10	95.4/9
	X0.25 Fo	98.5/4	85.2/3	73.5/10	98/5	89.8/10	73/6
	X0.5 Fo	99.5/6	88.3/10	88.3/10	98.5/3	91.3/5	88.8/10
	X0.75 Fo	100/4	89.3/6	96.4/10	99.5/4	90.8/9	79.1/5
	X1.0 Fo	99/3	92.9/10	100/10	100/5	94.4/10	96.9/10
	Best	10 Hz	15 Hz	X1.0	10 Hz	10 Hz	X1.0
Detect 2	Raw	99.5/8	81.9/5	71.3/6	99.5/8	81.9/5	71.3/6
	Frequency	98.5/5	81/4	67/5	99.5/5	83.5/7	70/8
	5 Hz	99/5	83/10	69.5/8	99/7	84.5/7	67.5/9
	10 Hz	99/8	83.5/5	73/4	100/3	90.5/6	84/10
	15 Hz	99/6	99/10	71/4	99.5/4	99/10	99.5/10
	X0.25 Fo	97/6	81/6	67/2	97.5/5	83.5/10	84/10
	X0.5 Fo	99/8	83/6	66.5/7	98/4	85.5/4	74.5/9
	X0.75 Fo	100/7	84.5/6	65/4	99/5	85/5	99.5/10
	X1.0 Fo	99/4	83/5	71.5/8	100/5	92.5/10	89/10
	Best	X.75	15 Hz	10 Hz	10 Hz	15 Hz	Several
Classify 1 vs. 2	Raw	99.5/8	83.7/5	68.9/9	99.5/8	83.7/5	68.9/9
	Frequency	99.5/5	86.7/6	73.5/4	100/4	89.3/7	73.5/6
	5 Hz	98/1	86.2/5	73.5/4	99.5/3	87.2/6	70.9/2
	10 Hz	99.5/2	86.7/2	74.5/4	99.5/2	88.3/3	73/6
	15 Hz	98.5/2	87.2/3	74.5/4	99/3	90.3/6	74/7
	X0.25 Fo	99/4	84.7/3	73.5/5	98.5/4	84.7/3	73.5/4
	X0.5 Fo	99/3	86.2/5	69.4/2	99.5/4	86.7/5	68.4/1
	X0.75 Fo	99/2	88.3/4	73.5/4	99/2	88.3/6	71.9/5
	X1.0 Fo	99.5/3	90.3/5	73/9	99.5/3	88.8/4	73.5/4
	Best	10 Hz	X1.0	Several	Freq.	15 Hz	15 Hz

Table A-5 Summary of Results for Artificial Data. Detection and Classification Accuracies in Percent.

Signals in Data Sets	SNR→	0 dB			−6 dB			−12 dB		
	Features→	Raw	Freq.	Filt.	Raw	Freq.	Filt.	Raw	Freq.	Filt.
Windowed Averaged EP	Detect 1	99	99.5	100	89.8	89.3	99	78.8	82.1	98.5
	Detect 2	99.1	99.5	100	92.5	93.5	100	83.5	85.5	99.5
	Classify	85.7	92.4	95.9	80.1	83.2	87.8	71.4	73	80.1
15 Gaussians	Detect 1	98.5	99	100	83.6	85.7	100	66.2	73	98
	Detect 2	96.1	100	100	78.5	81	98	70.6	74	98
	Classify	89.8	100	100	77.6	86.2	90.3	63.8	74	77
1 Windowed Sinusoid	Detect 1	98.1	100	100	84.7	92.9	100	73.4	79.1	99.5
	Detect 2	99	100	100	88.5	91	100	74.4	72.5	100
	Classify	99.5	100	100	83.7	95.9	97.5	71.9	78.1	82.7
3 Windowed Sinusoids	Detect 1	99.4	100	100	86.2	87.8	96.4	73.9	74	100
	Detect 2	99.5	99.5	100	81.9	83.5	99	71.3	70	99.5
	Classify	99.5	100	99.5	83.7	89.3	90.3	68.9	73.5	74.5

two signals at -6dB, raw features yielded an accuracy of 80.1% at 8 features. Frequency features yielded accuracies up to 83.2% at 5 features. Filtered features yielded accuracies to 87.8% at 7 features, which represents an 7.7% improvement over the raw feature accuracy.

Detection of these signals at SNR's of -12dB produced accuracies of 78.8% for 5 and 83.5% for 6 raw features, and 82.1% and 85.5% for 6 and 7 frequency features for the two data sets. Filtered features yielded significant improvements over these accuracies with 98.5% at 5 features and 99.5% at 8 features for the two data sets. These are improvements over raw feature accuracies of 19.7% and 16%. For classification between these two signals at SNR's of -12dB, 5 raw features yielded an accuracy of 71.4%, and frequency features yielded up to 73% accuracy at 3 features. Filtered features yielded up to 80.1% accuracy with 8 features. This is an improvement over the raw feature accuracy of 8.7%.

Signals which were composed of 15 Gaussian components, all with a width measured by the standard deviation of 15ms, represent complex signals that may be similar to EP's. Data containing these signals were tested and the results presented in Table A-2. At SNR levels of 0dB, averaged detection accuracies for raw features were 98.5% and 96.1% for 6 and 8 features for the two data sets. Frequency features yielded up to 99% and 100% at 6 features, and 4 filtered features produced 100% accuracies. Most bandwidths yielded similar accuracies. Classification between the two sets of data yielded an accuracy of 89.8% using 5 raw features. 100% was obtained using 4 frequency features. Accuracies of 100% were also achieved with filtered features, using only 2 features in some cases.

For detection of these Gaussian component signals in EEG at SNR's of -6dB, raw features yielded averaged accuracies of 83.6% for 5 features and 78.5% for 6 features for the two sets. Frequency features yielded 85.7% at 6 features and 81% at 7 features for the two sets, improvements over raw features of 2.1% and 2.5%. Filtered features produced improvements of up to 20.5% with accuracies of 100% and 98% achieved for 10 and 9 features for the two data sets. Classification between the two sets yielded 77.6% for 6 raw features, up to 86.2% for 6 frequency features, and 90.3% for 8 filtered features which represents a 12.7% improvement. The various bandwidths produced similar accuracies.

For detection of these Gaussian component signals with SNR's of -12dB, 6 raw features produced average accuracies of 66.2% and 70.6%. Frequency features improved on these results somewhat to 73% and 71.5% for 5 and 6

features. Filtered features yielded accuracies of 98% at 9 and 8 features for the two data sets. Classification between these data sets at SNR's of -12dB yielded 63.8% at 5 raw features, and frequency features yielded a 10.2% improvement to an accuracy of 74% for 8 features. Filtered features further improved these figures to 77% for 6 features, a 13.2% improvement in accuracy over the results using raw features.

For signals composed of windowed single sinusoids (Table A-3), the lower bound accuracies at 0dB were mostly 100% for the filtered amplitude features, while the raw amplitude features yielded accuracies of 98.1% and 99% for the two sets. Five raw features were selected to achieve this compared to 1 filtered feature selected to achieve 100% for several different bandwidths. Other bandwidths produced 100% accuracy with only 2 features. Two frequency features were selected to achieve 100% accuracy for frequency features without transformation. Classification accuracies were similar with 2 filtered features yielding 100% in some cases and 100% was achieved in most cases. Four raw features yielded 99.5% accuracy, and 100% for 3 frequency features was obtained.

For detection of the same signals in EEG at SNR's of -6dB, raw features yielded averaged accuracies of 84.7% and 88.5% with 7 features. Frequency features yielded up to 92.9% at 4 features. Filtered features produced improvements to 100% with 5 to 8 features for the two classes of data. For classification at -6dB, 83.7% at 10 raw features was achieved, and 8 frequency features yielded up to 95.9%. Filter features yielded accuracies of up to 97.5% with 7 features.

For detection of the same signals in EEG at SNR's of -12dB, raw features produced averaged accuracies of 73.4% and 74.4% for the two sets at 5 features. Accuracies for frequency features of up to 79.1% and 72.5% were achieved for the two sets at 5 and 7 features. Filtered features produced accuracies up to 100% at 9 features with 98 to 99% achieved in several other cases, representing improvements of up to 26% over the results for detection using only raw features. There is a small trend for higher accuracies achieved with the larger bandwidths. For classification at -12dB, raw features produced 71.9% accuracy with 3 features. Frequency features yielded improvements to 78.1% at 6 features. Filtered features yielded further improvements to 82.7% at 8 features, a 10.8% improvement over classification using only raw features. Generally the larger bandwidths yielded slightly higher classification accuracies.

For detection of signals composed of 3 windowed sinusoids in EEG (Table A-4) at SNR's of 0dB, raw features yielded 99.4% accuracy at 7 features, and

frequency features yielded up to 100% at 7 features and 99.5% at 5 features for the two data sets. Accuracies of 100% were achieved for many cases of filtered features with as little as 2 features. Classification of data with SNR's of 0dB produced accuracies of 99.5% for 8 raw features, up to 100% for 4 frequency features, and 99.5% for 2 filtered features.

For detection of the same signals in EEG with SNR's of -6dB, 86.2% and 81.9% averaged accuracies for 9 and 5 raw features were achieved for detection of the two data sets. Frequency features yielded similar accuracies, and filtered features achieved 100% with 10 features. There was a slight trend of larger bandwidths producing higher accuracies, and the best value of bandwidth was 15 Hz. Classification between these two data sets produced 83.7% accuracy for 5 raw features. Frequency features yielded up to 89.3% for 7 features, and filtered features increased the accuracy up to 90.3% for 5 features, representing a 6.6% improvement over the use of raw features alone.

For detection of the same signals in EEG at SNR's of -12dB, 73.9% and 71.3% average accuracies were achieved with 5 and 6 raw features. Frequency features yielded similar accuracies of up to 74% and 70% for 7 and 8 features. Filtered features yielded up to 100% using 10 features for the first set, and 99.5% using 10 features for the second set. Classification between the classes yielded 68.9% for 9 raw features. Four frequency features yielded improvements to 73.5% accuracy. Filtered features yielded 74.5% accuracy with 4 features.

Comparing the results for detection of the signals composed of 1 windowed sinusoid with those of the more complex signals composed of 3 windowed sinusoids, the accuracies achieved were quite similar at the various SNR levels. Raw features produced very similar accuracies, and filtered features improved the accuracies to almost 100% for both type signals at the 3 SNR levels. At SNR's of -12dB, the improvements in the accuracies for filtered features over raw features were very significant, ranging from 25.5% to 28.5%.

Appendix B: Recursive Algorithm for Mean-Square Error

A large part of the linear filter literature is dedicated to special computer algorithms which improve the speed of the calculations required to design the filters. When the matrix $\underline{\hat{R}}_{rr}$ is Toeplitz (as it is when the received data are stationary), an algorithm developed by Levinson (1946) is commonly used to efficiently invert the matrix. Morf (1977) has also described an efficient solution to the linear prediction equations. The second algorithm described in this chapter was developed by this author and is an efficient method for solving the robust matrix normal equations when the data are nonstationary. In order to begin, recall the dimensions of the matrices involved in Equation (4-18).

$$\begin{aligned} &(1)\ \underline{\hat{H}} \text{ is } N\times kN\,, \\ &(2)\ \underline{\hat{R}}_{s_d s} \text{ is } N\times kN\,, \\ &(3)\ \underline{\hat{R}}^{+}_{rr_t} \text{ is } kN\times kN\,, \text{ and} \\ &(4)\ \underline{P}_{rr} \text{ is } kN\times kN \end{aligned} \tag{B-1}$$

where N is the number of sample points in each data record and k is the number of channels. The spectral decomposition of $\underline{P}_{rr}$ is defined as

$$\underline{P}_{rr} = \sum_{i=1}^{kN} \beta_i \underline{\phi}_i \underline{\phi}_i^T \tag{B-2}$$

and the singular value decomposition of $\underline{\hat{R}}_{s_d s}$ as

$$\underline{\hat{R}}_{s_d s} = \sum_{i=1}^{N} \mu_i \underline{u}_i \underline{v}_i^T \tag{B-3}$$

where

$$\beta_1 \geq \beta_2 \geq \cdots \geq \beta_{kN} \tag{B-4}$$

and

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_N \geq 0. \tag{B-5}$$

The $\underline{\phi}_i$ are the unit norm eigenvectors of $\underline{P}_{rr}$. The $\underline{u}_i$ are the N-dimensional, unit norm eigenvectors of $\hat{\underline{R}}_{s_d s}\hat{\underline{R}}_{s_d s}^T$, and the $\underline{v}_i$ are the kN-dimensional, unit norm eigenvectors of $\hat{\underline{R}}_{s_d s}^T\hat{\underline{R}}_{s_d s}$.

In order to derive the recursive algorithm for computing the MNLS implementation, the three terms in Equation (4-18) are again considered separately. The first term is the $\mathrm{tr}\hat{\underline{R}}_{s_d s_d}$. This term is independent of l and is therefore constant. It is computed once and stored.

After inserting the matrix decompositions, the second term becomes

$$\mathrm{tr}\left[\sum_{i=1}^{N}\mu_i\underline{u}_i\underline{v}_i^T\sum_{j=1}^{l}\frac{1}{\lambda_i}\underline{q}_j\underline{q}_j^T\sum_{m=1}^{N}\mu_m\underline{v}_m\underline{u}_m^T\right] \tag{B-6}$$

Now the following definition is made:

$$v_{ij} = \underline{v}_i^T\underline{q}_j = \underline{q}_j^T\underline{v}_i \tag{B-7}$$

Inserting (B-7) into Equation (B-6) and rearranging a few terms gives

$$\mathrm{tr}\left[\sum_{i=1}^{N}\mu_i\sum_{j=1}^{l}\frac{v_{ij}}{\lambda_j}\sum_{m=1}^{N}\mu_m v_{mj}\underline{u}_i\underline{u}_m^T\right] \tag{B-8}$$

By noting that

$$\mathrm{tr}\left[\underline{u}_i\underline{u}_m^T\right] = \underline{u}_i^T\underline{u}_m = \begin{cases} 1, & i = m \\ 0, & i \neq m \end{cases} \tag{B-9}$$

Equation (B-8) simplifies to

$$\sum_{i=1}^{N}\mu_i^2\sum_{j=1}^{l}\frac{v_{ij}^2}{\lambda_j} = \sum_{j=1}^{l}\frac{1}{\lambda_j}\sum_{i=1}^{N}(\mu_i v_{ij})^2 \tag{B-10}$$

Equation (B-10) can be updated with a minimum number of calculations by evaluating v_{ij} for all i and j beforehand and using the first l-1 values to compute the lth value.

Consider now the third term

$$\mathrm{tr}\left[\sum_{i=1}^{N}\mu_i\underline{u}_i\underline{v}_i^T\sum_{j=1}^{l}\frac{1}{\lambda_j}\underline{q}_j\underline{q}_j^T\sum_{r=1}^{kN}\beta_r\underline{\phi}_r\underline{\phi}_r^T\sum_{s=1}^{l}\frac{1}{\lambda_s}\underline{q}_s\underline{q}_s^T\sum_{t=1}^{N}\mu_t\underline{v}_t\underline{u}_t^T\right] \tag{B-11}$$

Using Equation (B-9) and the following definition

$$d_{rj} = \underline{\phi}_r^T\underline{q}_j = \underline{q}_j^T\underline{\phi}_r \tag{B-12}$$

Equation (B-11) can be simplified to

$$\sum_{i=1}^{N}\mu_i^2\sum_{j=1}^{l}\frac{v_{ij}}{\lambda_j}\sum_{r=1}^{kN}\beta_r d_{rj}\sum_{s=1}^{l}\frac{d_{rs}}{\lambda_s}v_{is}. \tag{B-13}$$

The recursive algorithm for computing the mean-square error due to the third term vs. l can be derived inductively. Let $\varsigma_3^2(l)$ be the mean-square error due to the third term when keeping l eigenvalues. When keeping only the largest eigenvalue,

$$\varsigma_3^2(1) = \sum_{i=1}^{N}\mu_i^2\frac{v_{i1}}{\lambda_1}\sum_{r=1}^{kN}\beta_r d_{r1}\frac{d_{r1}}{\lambda_1}v_{i1}. \tag{B-14}$$

The mean-square error when using only the two largest eigenvalues is

$$\varsigma_3^2(2) = \sum_{i=1}^{N}\mu_i^2\left[\frac{v_{i1}}{\lambda_1}\sum_{r=1}^{kN}\beta_r d_{r1}\left\{\frac{d_{r1}}{\lambda_1}v_{i1} + \frac{d_{r2}}{\lambda_2}v_{i2}\right\}\right. \tag{B-15}$$
$$\left. + \frac{v_{i2}}{\lambda_2}\sum_{i=1}^{kN}\beta_r d_{r2}\left\{\frac{d_{r1}}{\lambda_1}v_{i1} + \frac{d_{r2}}{\lambda_2}v_{i2}\right\}\right]$$

This result can be written in terms of $\varsigma_3^2(1)$ as

$$\varsigma_3^2(2) = \varsigma_3^2(1) + \sum_{i=1}^{N}\mu_i^2\left[\frac{v_{i1}}{\lambda_1}\sum_{r=1}^{kN}\beta_r d_{r1}\frac{d_{r2}}{\lambda_2}v_{i2} + \frac{v_{i2}}{\lambda_2}\sum_{r=1}^{kN}\beta_r d_{r2}\sum_{s=1}^{2}\frac{d_{rs}}{\lambda_s}v_{is}\right] \tag{B-16}$$

This can be simplified by factoring $\frac{v_{i2}}{\lambda_2}$ out toward the left to obtain

$$\varsigma_3^2(2) = \varsigma_3^2(1) + \frac{1}{\lambda_2}\sum_{i=1}^{N}\mu_i^2 v_{i2}\sum_{r=1}^{kN}\beta_r d_{r2}\left[2\frac{d_{r1}}{\lambda_1}v_{i1} + \frac{d_{r2}}{\lambda_2}v_{i2}\right] \tag{B-17}$$

This can be extended to the general case when keeping l eigenvalues with the recursive algorithm

$$\varsigma_3^2(l) = \varsigma_3^2(l-1) + \frac{1}{\lambda_l}\sum_{i=1}^{N}\mu_i^2 v_{il}\sum_{r=1}^{kN}\beta_r d_{rl}\left[2\sum_{j=1}^{l-1}\frac{d_{rj}}{\lambda_j}v_{ij} + \frac{d_{rl}}{\lambda_l}v_{il}\right] \tag{B-18}$$

Equation (B-18) can be made even more efficient by recursively defining the summation from j=1 to l-1 as

$$\text{dsum}(l) = \text{dsum}(l-1) + \frac{d_{r(l-1)}}{\lambda_{l-1}}v_{i(l-1)} \tag{B-19}$$

where dsum(1) = 0. Using this definition, Equation (B-23) becomes

$$\varsigma_3^2(l) = \varsigma_3^2(l-1) + \frac{1}{\lambda_l}\sum_{i=1}^{N}\mu_i^2 v_{il}\sum_{r=1}^{kN}\beta_r d_{rl}\left[2\ \text{dsum}(l) + \frac{d_{rl}}{\lambda_l}v_{il}\right] \tag{B-20}$$

The general recursive algorithm for computing the mean-square error when keeping the largest l eigenvalues is given by combining the three foregoing terms.

$$\varsigma^2(l) = \varsigma^2(l-1) + \frac{1}{\lambda_l}\sum_{i=1}^{N}(\mu_i v_{il})^2 \tag{B-21}$$
$$+ \frac{1}{\lambda_l}\sum_{i=1}^{N}\mu_i^2 v_{il}\sum_{r=1}^{kN}\beta_r d_{rl}\left[2\ \mathrm{dsum}(l) + \frac{d_{rl}}{\lambda_l}v_{il}\right]$$

The initial values with which to start the recursion are

$$\varsigma^2(1) = \mathrm{tr}\left[\hat{\underline{R}}_{s_d s_d}\right] + \frac{1}{\lambda_1}\sum_{i=1}^{N}(\mu_i v_{i1})^2 \tag{B-22}$$
$$+ \frac{1}{\lambda_1^2}\sum_{i=1}^{N}(\mu_i v_{i1})^2\sum_{r=1}^{kN}\beta_r d_{r1}^2$$

and

$$\mathrm{dsum}(1) = 0. \tag{B-23}$$

By using Equation (B-21), the mean-square error can be evaluated for all values of l in order to determine l_{opt}. This l_{opt} can then be used to compute $\hat{\underline{R}}^{+}_{rr_s}$ in order to find the unique MNLS solution to the filter equations.

END

DTIC

8-86